CR-171731

(E84-10087) CROP AREA ESTIMATION BASED ON REMOTELY-SENSED DATA WITH AN ACCURATE BUT COSTLY SUBSAMPLE Final Report, 1 Jan. - 30 Oct. 1983 (Southern Methodist Univ.) 97 p HC A05/MF A01 CSCL 02C G3/43

N84-19958

Unclas 00087

CROP AREA ESTIMATION BASED ON REMOTELY-SENSED DATA WITH AN ACCURATE BUT COSTLY SUBSAMPLE

FINAL REPORT

ON

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

CONTRACT NO. NCC 9-9

1 January 1983 to 30 October 1983

Richard F. Gunst
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

TABLE OF CONTENTS

		Page
ABSTRACT		1
I.	RESEARCH ACTIVITIES	2
	A. ERRORS-IN-VARIABLES MODELS	3
	B. RELATED RESEARCH	5
II.	PROSPECTIVE FUTURE RESEARCH	6
III.	PRESENTATIONS AND OTHER ACTIVITIES	8
IV.	TITLES OF COMPLETED RESEARCH	9
V.	APPENDIX: COMPLETED MANUSCRIPTS	10

1. "Sensitivity of Errors-in-Variables Estimators to the Specification of the Ratio of Error Variances"
2. "Toward a Balanced Assessment of Collinearity Diagnostics"
3. "Regression Diagnostics and Approximate Inference Procedures for Penalized Least Squares Estimators"

ABSTRACT

This report documents research activities which were conducted from 1 January 1983 to 30 September 1983 under the auspices of National Aeronautics and Space Administration Contract NCC 9-9. During this contract period the primary focus of research was on alternatives to sampling-theory stratified and regression estimators of crop production and timber biomass. An alternative estimator which is viewed as especially promising is the errors-in-variable regression estimator. Investigations conducted during the course of this contract period established the need for caution with this estimator when the ratio of two error variances is not precisely known. One technical report on these investigations has been completed, a shorter version of which is being prepared for submission to a professional journal. In addition, further research topics on errors-in-variables estimation have been identified.

Richard F. Gunst
Principal Investigator
NASA Contract No. NCC 9-9

I. RESEARCH ACTIVITIES

Research supported by this contract is directed toward the study of estimators of crop production and timber biomass, two of potentially many applications. The specific focus of these investigations is on the development of methodologies which will enable satellite remote-sensing information to be combined with more accurate but more costly ground observation. Two of the classical estimation techniques for combining satellite data with ground truth are sampling-theory stratified estimation and regression estimation.

The assumptions which underly the use of the two sampling-theory estimators require that one of the sets of observations (satellite or ground-truth measurements) be known exactly, i.e., without measurement error. In some applications it is unreasonable to expect that either satellite or ground truth observations will be free from error. When this occurs a more appealing estimation methodology assumes an "errors-in-variables" regression model underlines the relationship between satellite and ground-truth measurements.

Section A below outlines the investigations which were conducted to assess the suitability of errors-in-variables models for application to the problems mentioned above. Section B describes related studies which were conducted on classical

regression estimators. It is anticipated that these related studies will be connected with errors-in-variables estimators in future investigations.

A. Errors-in-Variables Models

Let y denote a measurement (e.g., timber biomass) taken by ground observation and x a corresponding measurement obtained by satellite remote sensing. It is desirable to establish an empirical relationship between y and x so that the more readily obtained and less costly x measurements can be used to accurately estimate the more costly y values. Assuming that both y and x contain measurement error, a linear "errors-in-variables" regression model can be formulated as follows.

Denote the true (i.e., error free) ground-truth measurement by Y and the corresponding true satellite measurement by X. Assume that an adequate approximation to the relationship between Y and X is given by the linear model

$$Y = \alpha + \beta X.$$

In this setting Y and X cannot be observed because of measurement error; rather, one observes

$$y = Y + v \quad \text{and} \quad x = X + u,$$

where v and u are the measurement errors. In this framework the usual least squares estimators of α and β are biased since an underlying assumption for least squares estimation is that the predictor variable X is measured without error.

Most of the literature on errors-in-variables estimation is concerned with establishing conditions under which consistent estimators of α and β exist. If u, v, and X are assumed normally distributed and all model parameters are unknown, consistent estimators do not exist. If one or more of the model parameters are known (e.g., variances of the error measurements), consistent estimators of α and β are ordinarily available. In particular, if the ratio of the error variances, $\lambda = \sigma_v^2/\sigma_u^2$, is known then consistent estimators of α and β exist.

While the theoretical existence of consistent estimators of α and β has been an important topic of study, very little research has been conducted on (i) the effects of assuming an incorrect value for a model parameter and (ii) the construction of consistent estimators when X or u and v are assumed to be nonnormally distributed. A major achievement of the research conducted under this contract is an extensive investigation of the effects of assuming an incorrect value for the error variance ratio λ. The results of this investigation are reported in a manuscript entitled "Sensitivity of Errors-in-Variables Estimators to the Specification of the Ratio of Error Variances" which is appended to this report. In the near future this manuscript will be submitted for publication to a scientific journal.

The second research topic is currently being explored. The literature on errors-in-variables estimators established the existence of consistent estimators of α and β when X or (u,v) is nonnormal but no guidance is provided on how to construct consistent

estimators. Maximum likelihood estimation is generally intractable. Moment estimators exist and are consistent but moment estimation is known to be inefficient for finite sample sizes. An alternative approach which appears promising is outlined in Section II of this report.

B. Related Research

Least squares estimators are known to be seriously affected by the presence of outliers and collinearities, even if the requisite model assumptions are valid. The principal investigator has been actively investigating topics of importance to an understanding of outliers and collinearities over the past several years and is continuing to do so under this contract. It is anticipated that the results of these investigations will have an important impact on the application of errors-in-variables estimation, especially since the model framework admits the possible presence of outliers through the error terms.

Two manuscripts were completed on these topics during the current contract period. One manuscript is an invited critique of a manuscript on collinearity measures which will be published in the May 1984 issue of The American Statistician. The second manuscript presents new results on outlier diagnostics for ridge regression and smoothing spline estimators. This manuscript has been submitted for publication.

II. PROSPECTIVE FUTURE RESEARCH

As mentioned in the previous section, an important topic of research on errors-in-variables models is the construction of estimators when the true (unobservable) predictor variable X or the error terms u, v are not normally distributed. For example, in estimating crop proportions both y and x are bounded by the interval [0,1]. The study of theoretical properties of errors-in-variables estimators under the assumption that X follows a probability distribution over the unit interval (e.g., uniform or beta) would appear to be more reasonable than assuming an (unbounded) normal distribution.

Likelihood functions for (y,x) when X is nonnormal are generally theoretically intractable and fraught with computational difficulties. Moment estimators are easy to deal with but ordinarily inefficient for finite sample sizes. An alternative to maximum likelihood or moment estimation which is potentially fruitful for productive research and application is "pseudo maximum likelihood" estimation (e.g., Gong and Samaniego, Annals of Statistics, 1981). This theory allows all nuisance parameters to be replaced in the likelihood function by consistent estimators of the corresponding parameters and then the likelihood function is maximized with respect to the parameter(s) of interest. Future research will investigate asymptotic properties

of pseudo maximum likelihood estimators and compare their properties with moment and least squares estimators.

Once viable estimation methodologies are available under feasible model assumptions, errors-in-variables estimators will be compared with their sampling-theory counterparts. It is intended that both theoretical and empirical (using actual satellite and ground-truth measurements) comparisons will be conducted.

III. PRESENTATIONS AND OTHER ACTIVITIES

An oral presentation of preliminary results achieved under the support of this contract was made at a conference held at the Johnson Space Center in September, 1983. Further presentations are planned for future conference presentations at the Johnson Space Center. In addition, oral presentations are planned for national and regional meetings of the American Statistical Association, including the 1984 Annual Meetings next August in Philadelphia, PA.

During this contract period one advanced statistics graduate student in the Department of Statistics, Southern Methodist University,was supported by contract funds. Mani Y. Lakshminarayanan is currently conducting dissertation research on errors-in-variables models under the direction of the Principal Investigator. The investigations discussed in this report are a result of the collaboration between Mr. Lakshminarayanan and the Principal Investigator.

IV. TITLES OF COMPLETED RESEARCH

1. "Sensitivity of Errors-in-Variables Estimators to the Specification of the Ratio of Error Variances," revised manuscript under preparation for submission to a professional journal. (with M. Y. Lakshminarayanan).

2. "Toward a Balanced Assessment of Collinearity Diagnostics," The American Statistician, 38 (to appear, May 1984).

3. "Regression Diagnostics and Approximate Inference Procedures for Penalized Least Squares Estimators," submitted for publication. (with R. L. Eubank)

APPENDIX:

COMPLETED MANUSCRIPTS

SENSITIVITY OF ERRORS-IN-VARIABLES ESTIMATORS TO THE SPECIFICATION OF THE RATIO OF ERROR VARIANCES

Richard F. Gunst and Mani Y. Lakshminarayanan
Department of Statistics
Southern Methodist University
Dallas, Texas 75275

1. INTRODUCTION

Crop area estimation and the estimation of timber biomass are two applications of satellite remote sensing. Estimates obtainable with current technology often are not sufficiently precise for geographical regions which are as small as Crop Reporting Districts or counties. In order to improve the precision of these estimates based soley on remote-sensing information, field measurements are taken on relatively small portions of the geographical areas of interest. Stratified (sampling-theory) and regression estimation (e.g., Cochran 1963) are two statistical methodologies which can be used to combine the satellite information with that collected on the ground. In particular, regression estimation based on "errors-in-variables" (EV) models is viewed as an especially promising alternative for increasing the precision of satellite remote-sensing estimates of crop and biomass area. Most applications of EV estimation, however, require knowledge of a ratio of the variances of the measurement errors in order to obtain consistent estimates of the unknown regression coefficients. In this paper the sensitivity of EV estimators to the selection of the ratio of error variances is investigated.

Errors-in-variables models are appropriate when both variables in a regression model are subject to measurement error. Thus, a theoretical regression model specifying a relationship between a response variable (e.g., "ground-truth" crop or biomass area measurements) and a predictor variable (e.g., satellite area measurements) might be defined as follows:

$$Y = \alpha + \beta X + \varepsilon \ , \qquad (1.1)$$

where Y and X denote the true response- and predictor-variable values, respectively, and ε is an unknown model specification error. In practice, Y and X cannot be measured exactly; rather, one observes

$$x = X + u \quad \text{and} \quad y = Y + v \ , \qquad (1.2)$$

where u and v are measurement errors which may be correlated. Model (1.1) can now be expressed in terms of the observable quantities y and x as:

$$y = \alpha + \beta x + ([v+\varepsilon]-\beta u). \qquad (1.3)$$

Note that in model (1.3) the specification and measurement errors of the response variable (i.e., ε and v) are additive and are not separately estimable. Consequently, in this investigation the specification error is assumed zero or negligible relative to the measurement errors and the following reduced model is considered:

$$y = \quad + \beta x + (v-\beta u) \ ; \qquad (1.4)$$

equivalently, the EV model specification incorporates equations (1.1) and (1.2) with $\varepsilon = 0$.

Certain types of replication allow estimation of all EV

model parameters (see Kendall and Stuart 1977, Chapter 29). Likewise, measurement of additional variables which are correlated with the predictor variable X but not with the measurement errors can also allow consistent estimation of all model parameters (e.g., Durbin 1954; Feldstein 1974; Sargan 1958). Without either replication or the measuring of additional variates, it is possible to consistently estimate all EV model parameters only when one or more (functions of) the unknown parameters is known.

In Section 2 of this paper EV estimators are derived under normality assumptions and their lack of consistency is examined. The sensitivity of EV estimators when the ratio of the error variances is assumed known, the most prevalent side condition which is imposed to assure consistency, is examined in Section 3 by evaluating the derivative of the EV slope estimator under a variety of probabilistic assumptions on the unknown variance ratio. Section 4 presents a simulation study investigating the mean squared error properties of the EV slope estimator for a grid of assumed and true values of the unknown ratio of error variances. Concluding remarks are made in Section 5.

2. MAXIMUM LIKELIHOOD ESTIMATION

A distinction must be made between two assumptions about the true (unknown) response and predictor variables in model (1.1) before maximum likelihood estimators can be derived and their appropriateness evaluated. "Functional" EV models stipulate that the true underlying variates are constants whereas "structural" EV models assume that (Y,X) are realizations of some joint probability

distribution (e.g., Kendall and Stuart 1977, Chapter 29; Moran 1971). In this paper only the latter specification is studied; in particular, assume that

$$X \sim N(\mu,\sigma_X^2) \ , \tag{2.1}$$

which, from (1.1) (with $\varepsilon = 0$), necessarily implies normality for Y. In addition, it is common to assume that the measurement errors are jointly normally distributed, independently of (Y,X). Although correlation between the two measurement errors does not add substantive complexity, for simplicity and ease of presentation it is assumed that u and v are uncorrelated with

$$u \sim N(0,\sigma_u^2) \quad \text{and} \quad v \sim N(0,\sigma_v^2) \ . \tag{2.2}$$

Under these assumptions, differentiation of the likelihood function results in the following system of maximum likelihood estimating equations:

$$\begin{aligned} \bar{x} &= \hat{\mu} & \bar{y} &= \hat{\alpha} + \hat{\beta}\hat{\mu} \\ s_x^2 &= \hat{\sigma}_X^2 + \hat{\sigma}_u^2 & s_y^2 &= \hat{\beta}^2\hat{\sigma}_X^2 + \hat{\sigma}_v^2 \\ s_{xy} &= \hat{\beta}\hat{\sigma}_X^2 \end{aligned} \tag{2.3}$$

In equations (2.3) s_x^2 and s_y^2 are the sample variances of x and y, respectively, and s_{xy} is their sample covariance. There are six EV model parameters which must be obtained from these five estimating equations; equivalently, there are five sufficient statistics from which to estimate all six model parameters. Much theoretical work has been conducted to determine whether the six model parameters

are estimable under the normality assumptions (2.1) and (2.2). As will be detailed in Section 2.2, without additional knowledge about one or more of the model parameters it is impossible to consistently estimate all six parameters under the above normality assumptions. Before discussing the reasons for this lack of estimability, consider the solutions to equations (2.3) when one or both of the measurement error variances is known.

2.1 Estimation with Known Measurement Error Variances

Maximum likelihood estimates are the solutions to equations (2.3), provided that the solutions fall within the parameter space of the joint distribution of (X,u,v). Estimation of μ, α, and β pose no parameter space difficulties since the parameter space for each is the entire real line. Estimation of the variances σ_X^2, σ_u^2, and σ_v^2 requires that the solutions be nonnegative, leading to the following set of inequalities for the individual estimates:

$$
\begin{aligned}
&\text{For } \hat{\sigma}_u^2 \geq 0 && \text{(i)} && \hat{\beta} s_x^2 \geq s_{xy} \\
&\text{For } \hat{\sigma}_v^2 \geq 0 && \text{(ii)} && s_y^2 \geq \hat{\beta} s_{xy} \\
&\text{For } \hat{\sigma}_X^2 \geq 0 && \text{(iii)} && s_x^2 - \hat{\sigma}_u^2 \geq 0 \\
& && \text{(iv)} && s_y^2 - \hat{\sigma}_v^2 \geq 0 \\
& && \text{(v)} && \text{if } \sigma_X^2 > 0, \text{ sign } (\hat{\beta}) = \text{sign } (s_{xy}) \\
& && && \text{if } \sigma_X^2 = 0, \hat{\beta} \text{ is indeterminate.}
\end{aligned}
\tag{2.5}
$$

In the remainder of this paper it is assumed that solutions to equations (2.3) satisfy these inequalities; refer to Kendall and Stuart (1977) for alternatives when inequalities (2.5) are not satisfied.

Maximum likelihood solutions to equations (2.3) impose two implicit restrictions on the estimator of β:

$$\hat{\beta}(s_x^2 - \sigma_u^2) = s_{xy} \quad \text{and} \quad s_y^2 - \sigma_v^2 = \hat{\beta} s_{xy} \,, \tag{2.6}$$

with estimates replacing parameter values in (2.6) depending on which parameters are assumed known. Since both variances are nonnegative, equations (2.6) lead to the following inequality on the EV slope estimator:

$$|s_{xy}|/s_x^2 \leq |\hat{\beta}| \leq s_y^2/|s_{xy}| \,, \tag{2.7}$$

provided that $\hat{\beta}$, $s_{xy} \neq 0$ (which occurs with probability one). The lower limit in inequality (2.7) is the least squares slope estimate from the regression of y on x and is attained when it is known that no measurement error occurs with the predictor variable. The upper limit is the inverse of the least squares slope estimate from the regression of x on y. The upper limit is attained when it is known that no measurement error occurs with the response variable.

Several authors have attempted to circumvent the lack of estimability of β by defining estimators which are functions of the limits in inequality (2.7). Gini (1921) proposed the arithmetic mean of the limits. Teisser (1948) and Kaila (1980) suggested using the geometric mean of the limits. Pal (1980) showed that neither of these proposed estimators is consistent; moreover, he argued that any EV slope estimator which is intermediate to the two limits is optimal for some value of the ratio of the error variances but not optimal for others.

Since the main interest in EV model estimation centers on the relationship between the true response and predictor variables, estimation of α, β, μ, and σ_X^2 is of paramount importance. Thus some knowledge of the measurement error variances is required to solve equations (2.3). There are four special cases which can arise.

Case 1: σ_u^2 Known

In this case,

$$\hat{\beta} = s_{xy}/(s_x^2 - \sigma_u^2) \ . \tag{2.8}$$

Observe that if $\sigma_u^2 = 0$, this EV slope estimator is the usual least squares estimator and is equal (in magnitude) to the lower bound in (2.7).

Case 2: σ_v^2 Known

In this case,

$$\hat{\beta} = (s_y^2 - \sigma_v^2)/s_{xy} \ . \tag{2.9}$$

If $\sigma_v^2 = 0$, this EV slope estimator is the reciprocal of the least squares estimator from the regression of x on y and is equal (in magnitude) to the upper bound in (2.7).

Case 3: $\lambda = \sigma_v^2/\sigma_u^2$ Known

This assumption is the most frequently cited means of resolving the lack of a unique solution to equations (2.3). When this assumption is made, all the restrictions in (2.5) are satisfied unless $s_{xy} = 0$, which occurs with probability zero. In addition, this assumption does not require explicit knowledge of the exact

value of either of the measurement error variances, only the relative magnitude of the variances. Often it is reasonable to assume the measurement errors are of the same magnitude so that $\lambda = 1$. The resulting EV slope estimator is

$$\hat{\beta} = [(s_y^2 - \lambda s_x^2) + \{(s_y^2 - \lambda s_x^2) + 4\lambda s_{xy}^2\}^{1/2}]/2s_{xy}. \qquad (2.10)$$

Case 4: Both σ_u^2 and σ_v^2 Known

In this case, equations (2.3) result in two estimators of β; viz., estimators (2.8) and (2.9). Depending on whether equations (2.5) (iii) or (iv) are satisfied (with σ_u^2 and σ_v^2 replacing $\hat{\sigma}_u^2$ and $\hat{\sigma}_v^2$), $\hat{\beta}$ is either the solution to (2.10) or indeterminate (see Birch 1964).

2.2 Identifiability Under Normal Assumptions

The maximum likelihood estimating equations (2.3) are derived under the assumption that the predictor variable X and the measurement errors u and v are normally distributed, assumptions (2.1) and (2.2). Not only does this result in estimating equations which produce nonunique solutions, but also the parameter β is "nonidentifiable" in the joint distribution of (y,x). Identifiability is a distributional property which requires that only one set of parameters can give rise to any specific distribution of the observed random variables. Under assumptions (2.1) and (2.2), the joint distribution of (y,x) is bivariate normal with

$$\begin{aligned} \mu_y &= \alpha + \beta\mu & \mu_x &= \mu \\ \sigma_y^2 &= \beta^2\sigma_X^2 + \sigma_v^2 & \sigma_x^2 &= \sigma_X^2 + \sigma_u^2 \\ \sigma_{xy} &= \beta\sigma_X^2 \quad . \end{aligned} \qquad (2.11)$$

That this joint distribution is nonidentifiable can be demonstrated by the following sets of parameters from the distributions of (X,u,v), each of which produces a bivariate normal distribution for (y,x) with $\mu_y = \nu$, $\mu_X = \mu$, $\sigma_y^2 = \sigma_x^2 = 1$, and $\rho_{xy} = .5$ (Madansky 1959):

(a) $\sigma_X^2 = 1/2$, $\sigma_u^2 = 1/2$, $\sigma_v^2 = 1/2$, $\beta = 1$, $\alpha = \nu - \mu$

(b) $\sigma_X^2 = 1/3$, $\sigma_u^2 = 2/3$, $\sigma_v^2 = 1/4$, $\beta = 3/2$, $\alpha = \nu - 3\mu/2$.

Geary (1942) showed that when (u,v) are jointly normally distributed, if X possesses a finite nonzero cumulant of order greater than two then β is identifiable in the joint distribution of (y,x); thus, nonnormal distributions for X generally allow maximum likelihood estimation of β. Reiersol (1950) strengthened this result by proving that when the distribution of (u,v) is bivariate normal, nonnormality of X is a necessary and sufficient condition for identifiability of β. He also showed that if the distribution of X is normal, a necessary and sufficient condition for identifiability of β is that neither the distribution of u nor that of v is divisible by a normal distribution. Further, Reiersol established that once β is identifiable, so is α. If β is identifiable, he proved that a necessary and sufficient condition for identifiability of the other model parameters is that (i) the distribution of X (Y) is not divisible by a normal distribution and (ii) either u or v is identically zero. These important results on identifiability are summarized in Table 2.1

3. INFLUENCE OF λ ON THE EV SLOPE ESTIMATOR

The results of the previous section demonstrate that auxiliary knowledge must be available in order to estimate all six model parameters when normality of (X,u,v) is assumed. The most common assumption which is made is that the ratio of measurement errors λ is known. This allows estimation of β with assurance that the requisite restrictions (2.5) on the parameter estimates will hold. Likewise, this assumption does not require explicit knowledge of either of the measurement error variances.

Published research on EV model estimation has concentrated more on the existence of consistent estimators of β under various alternative assumptions than on the sensitivity of the resulting estimators to the assumptions. The dearth of sensitivity studies is surprising in light of the known lack of identifiability of β under the normality assumptions. The need for an evaluation of the sensitivity of equation (2.10) to the value of λ derives not only from the uncertainty of the robustness of the estimator to the choice of λ but also from parallel studies of other estimators which are similarly dependent on an unknown ratio of variances such as MINQUE variance component estimation. These latter studies (e.g., Hess 1979) have demonstrated that estimators which depend on selection of variance ratios can be affected by the choice of the ratio. No such guidance has been reported for EV parameter estimation.

Consider now the derivative of $\hat{\beta}$ with respect to λ. Asymp-

totically (i.e., replacing the sample moments by their corresponding parameter values),

$$\frac{\partial \hat{\beta}}{\partial \lambda} \to -\beta t/(\beta^2+\lambda), \tag{3.1}$$

where $t = \sigma_u^2/\sigma_X^2$ is the "noise-to-signal ratio" for the observable predictor variable x. From equation (3.1) one can readily see that the rate of change of $\hat{\beta}$ with respect to λ is not only a function of the value of λ but also of the true parameter value β and the noise-to-signal ratio. If the noise-to signal ratio is sufficiently small, equation (3.1) reveals that $\hat{\beta}$ will be relatively insensitive to the value of the true variance ratio λ. In addition, if for fixed t the true variance ratio λ is sufficiently large, $|(\partial\hat{\beta}/\partial\lambda)|/\beta$ will be relatively insensitive to the specific value of λ, especially if β is large.

That the estimator (2.10) can be extremely sensitive to the value of λ is illustrated in Figure 3.1. This figure graphs the (absolute) proportional rate of change of $\hat{\beta}$, $t/(\beta+\lambda)$, as a function of λ for two choices of the noise-to-signal ratio t and two choices of the true parameter β. The figure confirms that $\hat{\beta}$ is most sensitive to the choice of λ when t is large. For fixed t, the estimator is less sensitive to the choice of λ when λ is large, especially if coupled with a large β. In other words, under the precise conditions for which EV estimation is most often proposed (i.e., t moderate to large and λ small to moderate-each condition implying σ_u^2 is nonnegligible) the EV slope estimator is extremely sensitive to the true value of λ.

A somewhat different perspective on the sensitivity of $\hat{\beta}$ to the value of λ is obtained by assuming that λ is stochastic rather than deterministic. Lindley and El-Sayyad (1968) suggest using a Uniform (k^{-1},k) prior distribution for the error variance ratio if one believes that the two measurement errors are of the same magnitude. In addition, one might propose $N(k,\sigma_\lambda^2)$ or Chi-square(k) priors as reasonable alternatives in order to study the sensitivity of $\hat{\beta}$ to a variety of suspected prior distributions.

Given any of the above prior distributions for λ, one would like to evaluate the expected rate of change of $\hat{\beta}$ with respect to that prior; i.e., the expectation of (3.1) with respect to the prior on λ. Closed-form expectations do not ordinarily exist; however, the following theorem (e.g., Bishop, Feinberg, and Holland 1975, p. 493) allows approximate expectations to be determined.

Theorem 3.1 (Method of Statistical Differentials)

Let $g(x_1,x_2,\ldots,x_p)$ be a real-valued continuous function with continuous first and second derivatives at the point $\underline{\mu}=(\mu_1,\mu_2,\ldots,\mu_p)$. Let $\bar{x}_n = \{\bar{x}_{1n},\bar{x}_{2n},\ldots,\bar{x}_{pn}\}$ be a sequence of sample means of the vector random variable $\underline{x} = [x_1,x_2,\ldots,x_p]$. Finally, let $E[\underline{x}] = \underline{\mu}$ and let the distribution of $\underline{x}$ have finite third moments. Then

$$n^{1/2}[g(\bar{\underline{x}}_n) - g(\underline{\mu})] \rightarrow N(0,\Delta) ,$$

where

$$\Delta = [g^{(1)}(\underline{\mu}),\ldots,g^{(p)}(\underline{\mu})]\ddagger[g^{(1)}(\underline{\mu}),\ldots,g^{(p)}(\underline{\mu})]' ,$$

$g^{(i)}(\underline{\mu})$ is the partial derivative of $g(\underline{x})$ with respect to x_i evaluated at $\underline{x} = \underline{\mu}$, and $\ddagger$ is the variance-covariance matrix of $\underline{x}$.

Applying this theorem to the expectation of (3.1) under the three priors listed above for λ, the following approximations are obtained from a three-term Taylor expansion of $\partial\hat{\beta}/\partial\lambda$:

$\lambda \sim$ Uniform (k^{-1}, k)

$$\left| E\left[\frac{\partial\hat{\beta}}{\partial\lambda}\right]\right| = 2\beta t\{(2\beta^2+k+k^{-1})^{-1} + (k-k^{-1})^{-2}/3(2\beta^2+k+k^{-1})^3\} \qquad (3.2)$$

$\lambda \sim N(k, \sigma_\lambda^2)$

$$\left|E\left[\frac{\partial\hat{\beta}}{\partial\lambda}\right]\right| = \beta t\{(\beta^2+k)^{-1} + \sigma_\lambda^2(\beta^2+k)^{-3}\} \qquad (3.3)$$

$\lambda \sim$ Chi-Square (k)

$$\left| E\left[\frac{\partial\hat{\beta}}{\partial\lambda}\right]\right| = \beta t\{(\beta^2+k)^{-1} + 2k(\beta^2+k)^{-3}\} \; . \qquad (3.4)$$

Figures 3.2-3.4 depict the absolute proportional rate of change of $\hat{\beta}$, $|\partial\hat{\beta}/\partial\lambda|/\beta$, under the Uniform, Normal and Chisquare priors using equations (3.2)-(3.4). The Uniform prior displays the least sensitivity to the value of the variance ratio while both the Normal and the Chisquare priors produce large changes in $\hat{\beta}$, especially for small values of k. In each case the sensitivity is least when k is large and the noise-to-signal ratio t is small, as with the curves in Figure 3.1.

In each of Figures 3.2 to 3.4 the proportional rate of change of $\hat{\beta}$ is greatest when the parameter k of the prior distribution for λ is small. Thus if the measurement error in x tends to be much smaller than that of y and all other model parameters are fixed, the EV maximum likelihood estimator is relatively insensitive to

the exact value of λ; i.e., values of λ over a fairly wide range will result in similar, relatively small, estimator changes whereas small values of λ result in consequential estimator changes for the same values of other model parameters.

The results of this section can be summarized concisely as follows. First, if the measurement error in x is small relative to the measurement error in y and small relative to the variability in X (i.e., λ large, t small), then the EV maximum likelihood estimator of β, equation (2.10), will be relatively insensitive to the exact value of λ; therefore, one would expect that $\hat{\beta}$ would be relatively insensitive to erroneous selection of λ in a neighborhood of the true value. When the measurement error in x is large relative either to the error in y or to the variability in X, then the EV estimator of β is highly influenced by the true value of λ and, one would expect, to erroneous choices for λ in equation (2.10).

These results indicate that the selection of λ can be critical for accurate estimation of the slope parameter and one cannot merely assume that any "close" guess for the variance ratio will provide a suitable estimate. The simulation study reported in the next section documents more explicitly the dependence of the estimator on the correct choice of the measurement error variance ratio.

4. SIMULATION STUDY

Asymptotic properties of EV model estimators are often cited with little regard to whether they are valid for finite sample

sizes. In particular, asymptotic variance formulae are used to compare alternative estimators and to draw inferences on model parameters. In this section results of a simulation study are examined in order to (i) determine whether asymptotic variance formulae are adequate approximations to the true variances for finite samples, (ii) gauge the magnitude of the effects of assuming an incorrect value for the measurement error variance ratio λ, and (iii) assess the relative merits of least squares and EV estimators.

The following simulation results fix the values of β, σ_u^2, and σ_X^2 at 3.0, 5.0, and 5.0, respectively, so that by varying σ_v^2 the results are only a function of λ. Under the assumption of a known (correct) variance ratio, the EV estimator of β, equation (2.10) is asymptotically unbiased; i.e.,

$$\text{plim}(\hat{\beta}) = \beta \ . \tag{4.1}$$

The asymptotic variance of $\hat{\beta}$ is well known (e.g., Robertson 1974; Gleser 1981):

$$\text{asvar}(\hat{\beta}) = n^{-1}\{(\beta^2+\lambda)t + \lambda t^2\} \ . \tag{4.2}$$

In Tables 4.1 and 4.2, the mean and mean squared error of $\hat{\beta}$ are compared to the asymptotic values (4.1) and (4.2) for N=1,000 samples of size 20, 50 and 100. The ratios tabulated in Table 4.1 are the sample means of the 1,000 $\hat{\beta}$ values divided by β. In Table 4.2, the sample mean squared error, $\Sigma(\hat{\beta}_i-\beta)^2/N$, of $\hat{\beta}$ is compared to the corresponding theoretical values calculated from (4.2). In all cases, the sample means and mean squared errors in Tables 4.1 and 4.2 are evaluated using the assumed values of

λ, and compared to equations (4.1) and (4.2) using the true value of λ. In this way both the effect of sample size and the effect of an incorrect choice of λ can be assessed.

The entries in Table 4.1 corresponding to the correct assumed values for λ indicate close agreement between the average estimated EV slope values and the true parameter values, especially when the sample size is at least 50. In Table 4.2 agreement is not as good for small sample sizes but appears to be adequate for samples of size 100 if λ not too large. It would appear that a sample of size 200 would be adequate for acceptable agreement between the sample variance and the asymptotic variance. Tables 4.3 and 4.4 exhibit comparisons of sample mean squared errors to asymptotic variances for 1,000 samples of size 200 for three values of β and two choices of the noise-to-signal ratio t. In most of the model configurations the agreement is quite adequate when the correct value of the variance ratio is assumed, especially when β is large and t is small.

The off-diagonal elements of Tables 4.1 to 4.4 reveal the effects of incorrectly guessing the variance ratio λ. Incorrectly assuming too large a value of λ results in underestimation of $\hat{\beta}$ while the reverse is true when λ is assumed too small (Table 4.1). Any incorrect guess for λ produces an over-estimate (except for a few ratios which are within sampling error) for the asymptotic mean squared error of $\hat{\beta}$, but it is far more serious to guess too small a value for λ than too large a one (Tables 4.2-4.4).

Small sample sizes yield very erratic results; when n=20, the small sample estimates are unreliable as measures of asymptotic mean squared error (recall that the EV estimator is asymptotically unbiased). Even with larger sample sizes the sample mean squared errors are only reliable estimates of the asymptotic variances when the assumed value of λ is in a narrow interval around the true error variance ratio. For samples of size 200, the agreement between sample and asymptotic mean squared errors is adequate if the assumed λ is between approximately half and double the true ratio, especially--as the results of the previous section suggested--when the true value of λ is large, the noise-to-signal ratio is small, and β is large. While not confirmatory, this empirical finding about the close agreement between sample and asymptotic mean squared errors supports the use of Lindley and El-Sayyad's (1968) uniform prior in studies of the properties of EV slope estimators.

Another comparison which is of importance is that of the mean squared error of the EV maximum likelihood estimator to that of the least squares estimator, $mse(\hat{\beta})/mse(\hat{\beta}_{LS})$. Table 4.5 displays the ratios of the sample mean squared errors for the two estimators using the same model configurations as in Tables 4.1 and 4.2. It is evident from this table that $\hat{\beta}$ offers substantial improvement over least squares unless the sample size is small or the assumed value of λ is much less than the true value. As the sample size increases, only assumed values of λ which are grossly smaller than the true ones will lead to a preference for least squares over EV

estimation.

Lest these conclusions be affected by the inadequacy of empirical mean squared errors as estimates of asymptotic mean squared errors, Table 4.6 displays the ratios of the asymptotic mean squared errors of the EV slope estimators to the corresponding ones for least squares. The asymptotic mean squared errors for the EV estimators are based on assuming an incorrect value λ^* for λ. The appropriate expressions for the two estimators are:

$$\text{asmse}(\hat{\beta}_{LS}) = \beta^2 t(t+n^{-1})(1+t)^{-2} + n^{-1}\lambda(1+t^{-1})^{-1} \qquad (4.3)$$

and

$$\begin{aligned}\text{asmse}(\hat{\beta}) = {} & [-(\beta^2+\lambda^*)+(\lambda-\lambda^*)t + \{[\beta^2-\lambda^*)+(\lambda-\lambda^*)t]^2+4\lambda^*\beta^2\}^{1/2}]^2/4\beta^2 \\ & + \{3\beta^2t^2(\lambda-\lambda^*)^2+(\beta^4-4\lambda^*\beta^2+\lambda^{*2})[3\beta^2+(\lambda+\beta^2)t + \lambda t^2] \\ & - 3\beta^2(\beta^2-\lambda^*)^2 + 6\lambda^*\beta^2[\beta^2+(\lambda+\beta^2)t+\lambda t^2]\}/n(\beta^2+\lambda^*)^2. \qquad (4.4)\end{aligned}$$

As is apparent from Table 4.6, the conclusions drawn from Table 4.5 remain valid when the sample mean squared errors are replaced by asymptotic mean squared error.

Finally, Figures 4.1 to 4.6 display asymptotic mean squared error comparisons of $\hat{\beta}$ and using equations (4.3) and (4.2) or (4.4), as appropriate. The horizontal axes in Figures 4.1 and 4.2 are the variance proportions $\sigma_u^2/(\sigma_X^2+\sigma_u^2) = t/(1+t)$, a monotonic transformation of the noise-to-signal ratio. As indicated by Figures 4.1 and 4.2, for a fixed sample size and a fixed variance proportion the EV estimator (2.10) using the correct value of λ is preferable to least

squares only for small values of λ (recall, small λ implies the error in y is comparable or less than the error in x). As the noise-to-signal ratio decreases, the range of values of the variance ratio λ for which $\hat{\beta}$ offers improvement over least squares increases. Increasing the sample size enlarges the (t,λ)-region for which EV estimators are preferable to least squares.

Figure 4.3 graphs the ratios of $\text{asmse}(\hat{\beta})$ to $\text{asmse}(\hat{\beta}_{LS})$ using equations (4.2) and (4.3), respectively, for four sample sizes and $(\beta,\sigma_x^2,\sigma_v^2) = (3,5,10)$. Except for extremely small values of λ, the ratio of asymptotic mean squared errors is less than unity, especially so for large sample sizes. These graphs confirm the conclusions drawn from Tables 4.5 and 4.6 for correct choices of λ.

Figures 4.4 to 4.6 graph the ratio of the asymptotic mean squared errors using equation (4.4) for EV estimation with an incorrect choice for λ. Model parameters for these figures are $(\beta,\sigma_x^2) = (3,5)$ and $\lambda = 1$, 6 and 10, respectively. Again the conclusions drawn from Tables 4.5 and 4.6 are graphically confirmed from these figures: unless λ is selected much smaller than its true value the EV estimator of β is preferable to least squares.

5. CONCLUDING REMARKS

The results of Sections 3 and 4 establish the extreme sensitivity of the EV maximum likelihood estimator to the choice of the measurement error variance ratio. The sensitivity is model dependent and is greater when the true variance ratio λ is small and

the noise-to-signal ratio t is large. Large sample sizes enable the EV estimator to offer improvement over least squares if the assumed measurement error variance ratio is not too much smaller than the true ratio; however, the reliable use of asymptotic formulae for estimator variances requires that the variance ratio be known within a narrow interval of the true value and that the sample size be at least 200.

Throughout Sections 3 and 4 the EV slope estimator shows least sensitivity to the choice of the variance ratio λ when the noise-to-signal ratio $t = \sigma_u^2/\sigma_X^2$ is small and $\lambda = \sigma_v^2/\sigma_\iota^2$ is large, especially for large values of β. Together the conditions on λ and t imply that there is relatively little error in the predictor variable (i.e., $\sigma_u^2 \approx 0$). Thus, the model configurations for which the EV slope estimator is relatively insensitive to the choice of λ are those for which least squares is most appropriate. In other model configurations (i.e., when the error in x is not negligible) the EV slope estimator exhibits demonstrable sensitivity to the assumed value of the measurement error variance ratio.

In spite of these limitations on the application of EV estimation, the simulation results and asymptotic mean squared error comparisons in Section 4 indicate clear preference for EV estimation over least squares. If sample sizes are at least 200, this general conclusion is violated only when the assumed value of λ is much less than the true value.

Little insight can be gained from this study relative to the performance of EV maximum likelihood estimators under non-

normal assumptions. While the parameters become identifiable under the conditions stated in Table 2.1, analytic derivations of estimators and asymptotic variances are intractable for most alternatives to the normality assumptions. This important area of research is currently under investigation.

6. ACKNOWLEDGMENT

This research was supported by the National Aeronautics and Space Administration under Contract NCC 9-9.

REFERENCES

1. Birch, M. W. (1964). "A Note on the Maximum Likelihood Estimation of a Linear Structural Relationship," Journal of the American Statistical Association, 59, 1175-1178.

2. Bishop, Fienberg, Holland. (1975) Discrete Multivariate Analysis: Theory and Practice, MIT Press.

3. Cochran, W. G. (1977). Sampling Techniques , Third Edition, New York: John Wiley and Sons, Inc.

4. Durbin, J. (1954). "The Errors-in-Variables," International Statistical Review, 22, 23-32.

5. Feldstein, M. (1974). "Errors-in-Variables: Consistent Estimates with Smaller MSE in Finite Samples," Journal of the American Statistical Association, 69, 990-996.

6. Geary, R. C. (1942). "Inherent Relations Between Random Variables," Proceedings of Royal Irish Academy, 47, 63-76.

7. Gini, C. (1921). "Sull' interpolazione di una retta quando i valori della variabile independente sono affetti da errori accidentalli," Metron, 1, 63-82.

8. Gleser, L. J. (1981). "Estimation in a Multivariate EV Regression Model: Large Sample Results," Annals of Statistics, 9, 24-44.

9. Hess, J. L. (1979): "Sensitivity of MINQUE with Respect to a Priori Weights," Biometrics, 35, 645-650.

10. Kaila, K. L. (1980). "A New Method for Linear Curve Fitting with Both Variables in Error," Sankhya, Series B, 42, 81-97.

11. Kendall, M.G. and Stuart, A. (1977). The Advanced Theory of Statistics, Fourth Edition, Vol. 2, Macmillan Co.: New York.

12. Lindley, D.V. and El-Sayyad, G.M. (1968). "The Bayesian Estimation of a Linear Functional Relationship," Journal of the Royal Statistical Society, Series B, 30, 190-202.

13. Madansky, A. (1959). "The Fitting of Straight Lines When Both Variables are Subject to Error," Journal of the American Statistical Association, 54, 173-205.

14. Moran, P.A.P. (1971). "Estimating Structural and Functional Relationships," Journal of Multivariate Analysis, 1, 232-255.

15. Pal, M. (1980). "Some Observations on Estimation of Bivariate Relationships in the presence of Errors-in-Variables," Sankhya, Series B, 42, 98-109.

16. Reiersol, O. (1950). "Identifiability of a Linear Relation Between Variables which are Subject to Error," Econometrica, 18, 375-389.

17. Robertson, C. (1974). "Large Sample Theory for Linear Structural Relations," Biometrika, 61, 353-359.

18. Sargan, J. D. (1958). "The Estimation of Econometric Relationships Using Instrumental Variables," Econometrica, 26, 393-415.

19. Teisser, G. (1948). "La relation D'allometric sa Signification Statistique et Biologigue," Biometrics, 4, 14-53.

Table 2.1 Identifiability Conditions for EV Models*

(a) Identifiability of β

(i) If (u,v) is Normal, then X cannot be Normally Distributed

(ii) If X is Normal, the distribution of neither u nor v can be divisible by a Normal Distribution

(b) β is Identifiable

(i) α is Identifiable

(ii) All other Model Parameters are Identifiable iff

(1) The distribution of X (Y) is not divisible by a Normal Distribution, and

(2) Either $u = 0$ or $v = 0$

---

*All model parameters unknown.

Table 4.1: Ratio of Simulated and Asymptotic Expectations of EV Slope Estimator

	Assumed λ								
True λ	0.2	0.5	1.0	1.5	2.0	4.0	6.0	8.0	10.0
			(a) n = 20						
0.2	1.16	1.04	0.99	0.95	0.89	0.77	0.70	0.65	0.62
1.0	1.21	1.13	0.92	1.02	0.99	0.81	0.73	0.68	0.66
2.0	1.40	1.29	1.15	1.09	1.09	0.90	0.79	0.73	0.69
10.0	2.49	2.08	1.93	2.20	2.30	2.26	1.43	1.25	1.10
			(b) n = 50						
0.2	1.02	0.99	0.94	0.90	0.86	0.74	0.68	0.64	0.62
1.0	1.11	1.09	1.02	0.98	0.93	0.81	0.72	0.68	0.64
2.0	1.23	1.19	1.14	1.08	1.02	0.87	0.78	0.71	0.67
10.0	2.27	2.19	2.13	2.04	1.98	1.62	1.37	1.19	1.04
			(c) n = 100						
0.2	1.01	0.98	0.93	0.89	0.85	0.74	0.67	0.64	0.61
1.0	1.10	1.07	1.01	0.96	0.92	0.79	0.72	0.67	0.64
2.0	1.22	1.17	1.12	1.07	1.02	0.87	0.77	0.71	0.67
10.0	2.16	2.12	2.03	1.93	1.85	1.56	1.33	1.14	1.03

Table 4.2: Ratio of Simulated and Asymptotic Mean Squared Errors of EV Slope Estimators

	Assumed λ								
True λ	0.2	0.5	1.0	1.5	2.0	4.0	6.0	8.0	10.0
				(a) n = 20					
0.2	108.11	2.69	1.94	2.02	1.31	2.48	2.68	2.83	3.23
1.0	9.91	3.32	550.84	2.72	1.99	1.66	1.81	2.29	2.68
2.0	35.78	10.48	80.04	90.50	3.42	2.38	2.24	1.80	2.24
10.0	4553.13	996.18	3087.32	5074.26	357.29	1563.05	43.71	92.26	25.27
				(b) n = 50					
0.2	1.27	1.15	1.20	1.35	1.86	3.81	5.29	6.54	7.40
1.0	2.21	1.92	1.53	1.13	1.25	2.25	3.69	4.72	5.70
2.0	3.79	3.08	2.35	1.81	1.29	1.44	2.37	3.45	4.18
10.0	36.15	30.49	30.84	42.32	26.79	18.65	4.80	4.64	1.69
				(c) n = 100					
0.2	1.19	0.99	1.37	1.94	3.07	6.89	10.42	13.12	14.69
1.0	2.27	1.63	1.15	1.16	1.44	4.28	7.04	9.35	11.00
2.0	5.07	3.69	2.29	1.61	1.15	2.06	4.33	6.37	7.98
10.0	48.73	46.22	38.95	31.49	27.20	12.54	5.58	2.02	1.29

Table 4.3: Ratio of Simulated and Asymptotic Mean Squared Errors of EV Slope Estimator: n = 200, t = 0.1

True λ	Assumed λ								
	0.2	0.5	1.0	1.5	2.0	4.0	6.0	8.0	10.0
				(a) β = 0.1					
0.2	1.10	1.23	1.44	1.48	1.38	1.57	1.49	1.54	1.62
1.0	9.31	1.40	1.10	1.08	0.97	1.08	0.98	1.01	1.06
2.0	118.04	3.44	1.38	1.14	0.98	0.97	1.00	1.03	0.94
10.0	3046167.58	2823896.24	6151.68	9.22	3.85	1.55	1.08	1.04	0.98
				(b) β = 1.0					
0.2	1.08	1.65	3.49	5.09	6.35	9.13	10.93	11.47	12.11
1.0	5.45	2.30	1.06	1.46	1.88	4.13	5.28	6.15	6.45
2.0	16.43	7.97	2.90	1.35	1.07	1.98	2.85	3.58	3.87
10.0	137.29	95.94	53.21	31.12	18.68	4.15	1.95	1.20	1.06
				(c) β = 10.0					
0.2	1.03	1.06	1.03	1.06	1.04	1.07	1.10	1.09	1.03
1.0	1.06	1.03	1.03	1.04	1.04	1.06	1.01	1.13	1.15
2.0	1.07	1.07	1.09	0.95	1.02	1.07	1.02	1.07	1.15
10.0	1.26	1.22	1.23	1.08	1.22	1.16	1.13	1.13	1.10

Table 4.4: Ratio of Simulated and Asymptotic Mean Squared Errors of EV Slope Estimator, n = 200, t = 1.0

True λ	Assumed λ								
	0.2	0.5	1.0	1.5	2.0	4.0	6.0	8.0	10.0
				(a) β = 0.1					
0.2	1.08	0.99	1.18	1.31	1.33	1.40	1.48	1.45	1.42
1.0	1.57	31592.99	1.08	0.67	0.56	0.51	0.51	0.53	0.50
2.0	11789245.73	238241.92	57182.60	3.27	1.10	0.51	0.44	0.41	0.39
10.0	38414594.50	4864779.56	2293212.56	795824.42	703991.66	89744.73	694.63	2.28	1.12
				(b) β = 1.0					
0.2	1.03	5.62	15.38	21.15	24.16	29.89	31.91	32.85	33.34
1.0	39.90	11.69	1.17	2.76	5.10	10.67	12.91	13.71	14.54
2.0	128.86	72.34	19.83	4.11	1.19	3.97	6.00	7.16	7.46
10.0	12090.97	11811.55	31704.10	136241.72	837.30	227.65	17.88	2.83	1.12
				(c) β = 10.0					
0.2	1.01	1.09	1.05	1.13	1.02	1.20	1.49	1.82	2.28
1.0	1.13	1.13	1.10	1.12	1.06	1.17	1.30	1.65	2.21
2.0	1.25	1.21	1.07	1.11	1.11	1.15	1.20	1.52	1.91
10.0	2.67	2.83	2.76	2.61	2.48	1.91	1.47	1.25	1.11

Table 4.5: Ratio of Simulated Mean Square Error of EV and Least Squares Estimators

True λ	Assumed λ 0.2	0.5	1.0	1.5	2.0	4.0	6.0	8.0	10.0
				(a) n = 20					
0.2	21.13	0.53	0.38	0.40	0.26	0.47	0.51	0.55	0.63
1.0	2.26	0.78	127.43	0.62	0.46	0.37	0.42	0.52	0.63
2.0	9.73	2.83	20.96	23.92	0.92	0.64	0.61	0.48	0.62
10.0	2428.26	538.87	1683.58	2776.03	190.44	844.80	24.59	50.14	14.04
				(b) n = 50					
0.2	0.12	0.09	0.11	0.12	0.16	0.31	0.44	0.53	0.62
1.0	0.21	0.17	0.13	0.12	0.11	0.23	0.36	0.46	0.55
2.0	0.45	0.26	0.29	0.20	0.15	0.16	0.27	0.39	0.47
10.0	10.82	7.43	6.59	5.94	3.48	1.62	1.15	1.19	1.53
				(c) n = 100					
0.2	0.05	0.04	0.06	0.08	0.13	0.28	0.43	0.54	0.61
1.0	0.11	0.08	0.06	0.06	0.07	0.21	0.34	0.45	0.52
2.0	0.29	0.21	0.13	0.09	0.07	0.12	0.25	0.36	0.46
10.0	6.09	5.73	4.87	3.98	3.38	1.57	0.71	0.25	0.16

Table 4.6: Comparison of Asymptotic Mean Squared Errors of Least Squares and EV Slope Estimators

True λ	Assumed λ 0.2	0.5	1.0	1.5	2.0	4.0	6.0	8.0	10.0
				(a) n = 20					
0.2	0.19	0.20	0.23	0.26	0.31	0.51	0.68	0.82	0.92
1.0	0.26	0.24	0.23	0.24	0.27	0.43	0.60	0.75	0.86
2.0	0.43	0.38	0.31	0.28	0.27	0.36	0.52	0.66	0.78
10.0	5.14	4.71	4.06	3.48	2.98	1.58	0.89	0.62	0.56
				(b) n = 50					
0.2	0.08	0.09	0.11	0.14	0.19	0.37	0.53	0.65	0.73
1.0	0.13	0.11	0.10	0.13	0.13	0.29	0.44	0.57	0.67
2.0	0.27	0.22	0.15	0.12	0.11	0.20	0.35	0.47	0.59
10.0	4.86	4.43	3.79	3.21	2.71	1.29	0.58	0.30	0.24
				(c) n = 100					
0.2	0.04	0.05	0.07	0.10	0.14	0.33	0.48	0.59	0.67
1.0	0.08	0.06	0.05	0.06	0.08	0.23	0.39	0.51	0.60
2.0	0.21	0.16	0.10	0.07	0.06	0.14	0.29	0.41	0.52
10.0	4.75	4.33	3.69	3.11	2.61	1.18	0.47	0.19	0.12

Fig 3.1: Proportional Rate of Change of EV maximum likelihood estimator versus lambda

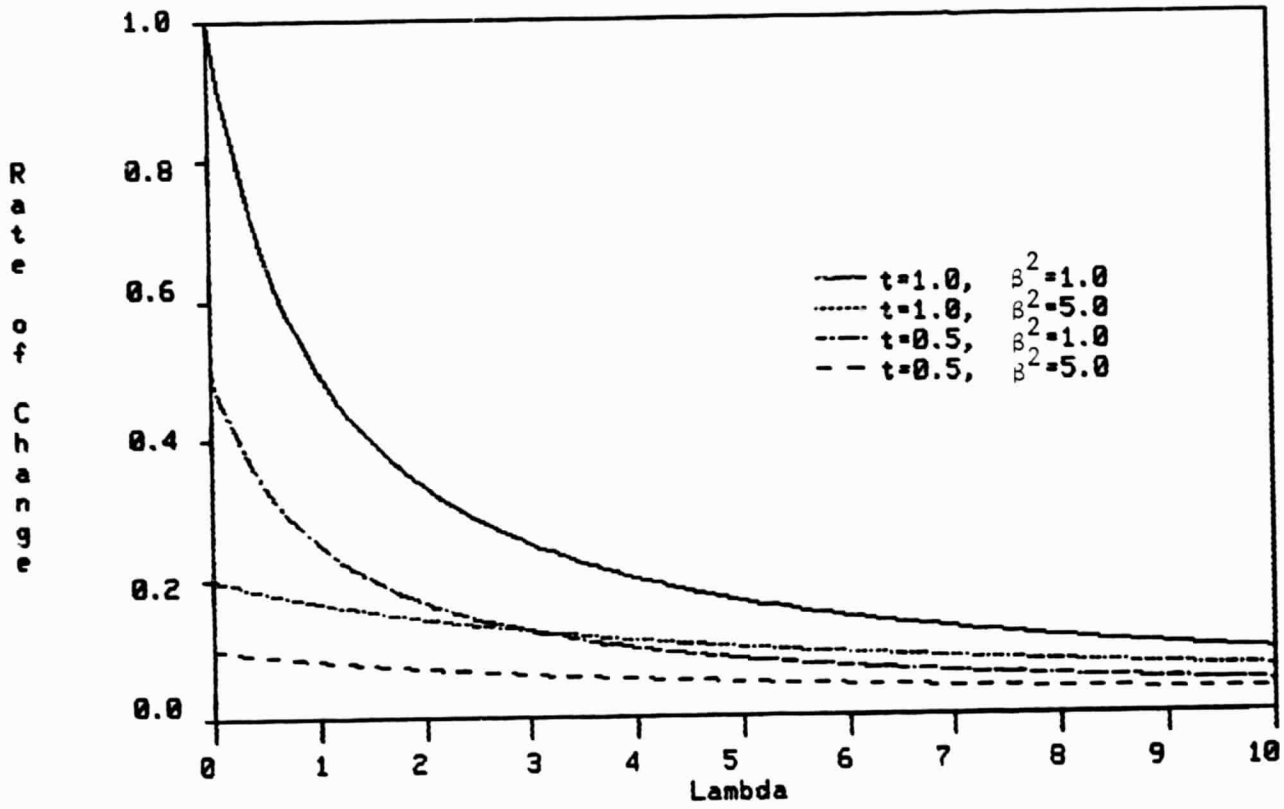


Fig 3.2: Expected Rate of Estimator Change
with Uniform Prior

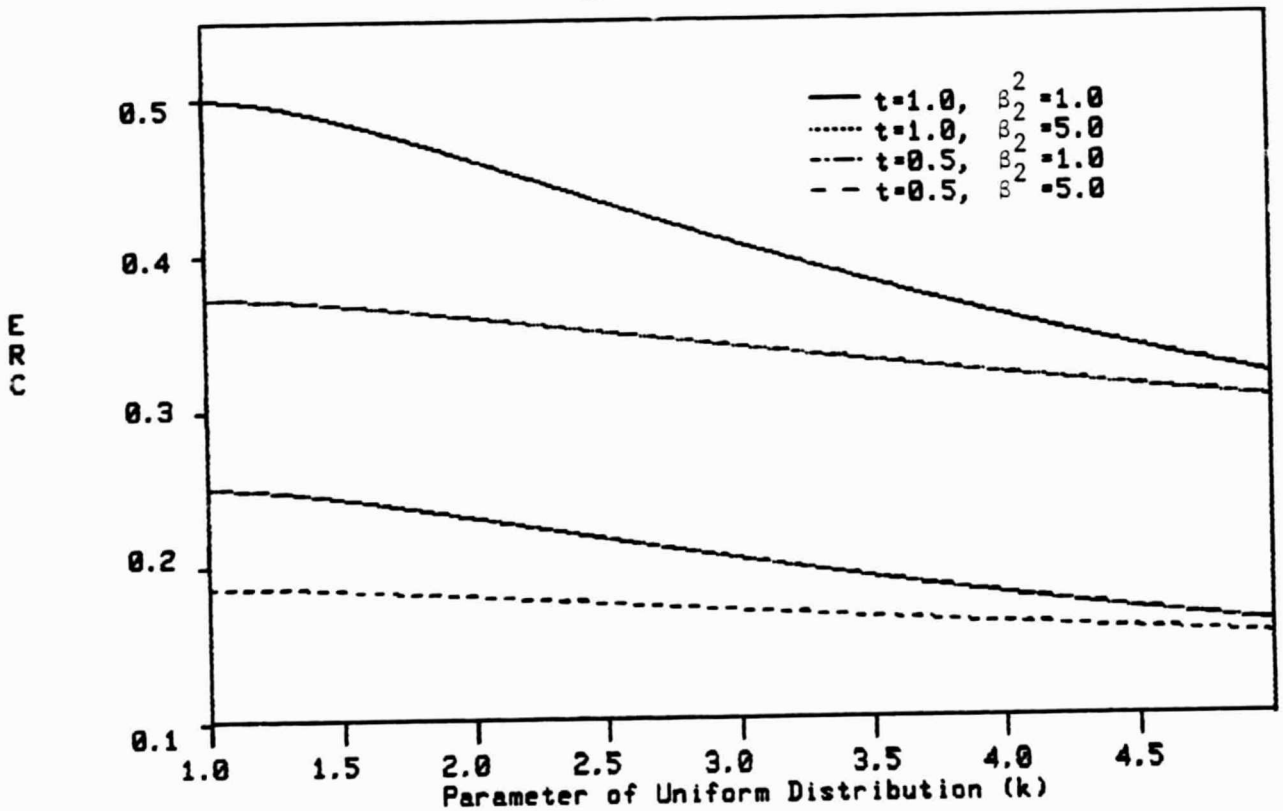


Fig 3.3: Expected Rate of Estimator Change with Normal Prior

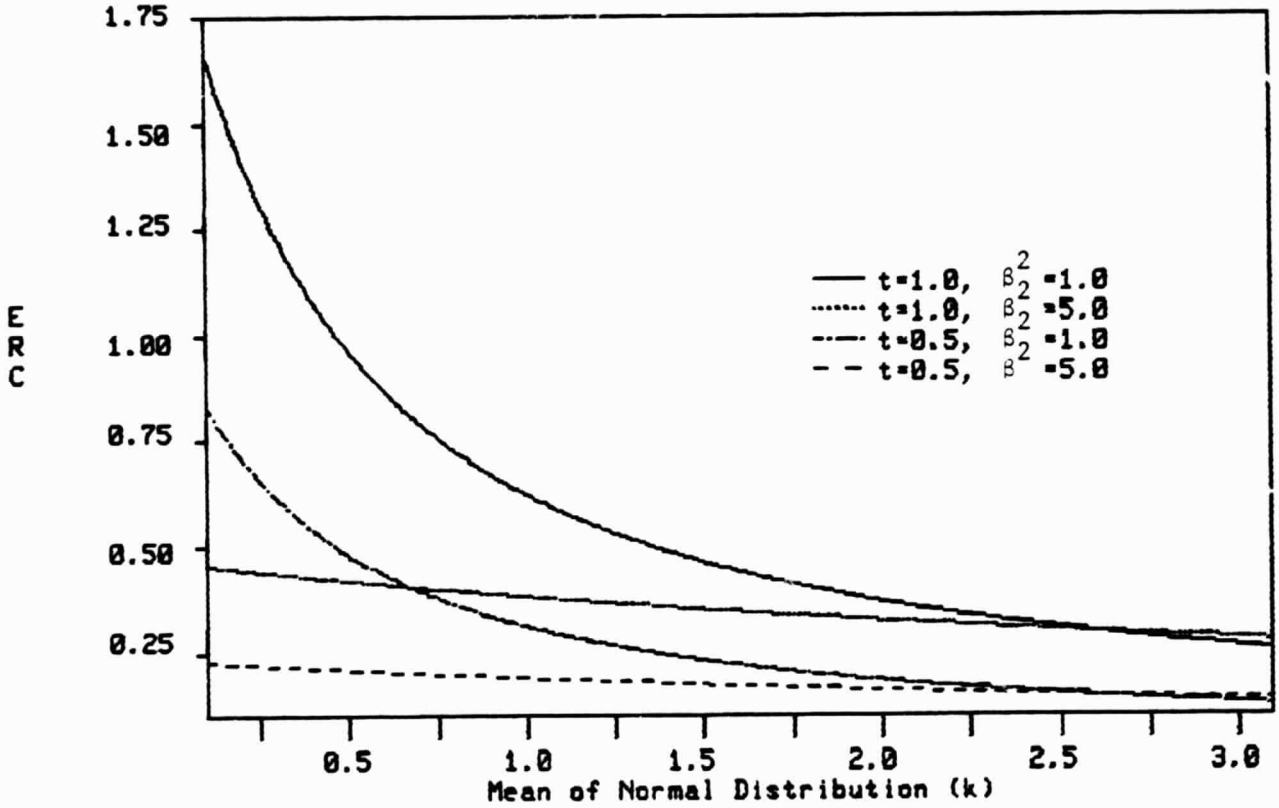


Fig 3.4: Expected Rate of Estimator Change
with Chisquare Prior

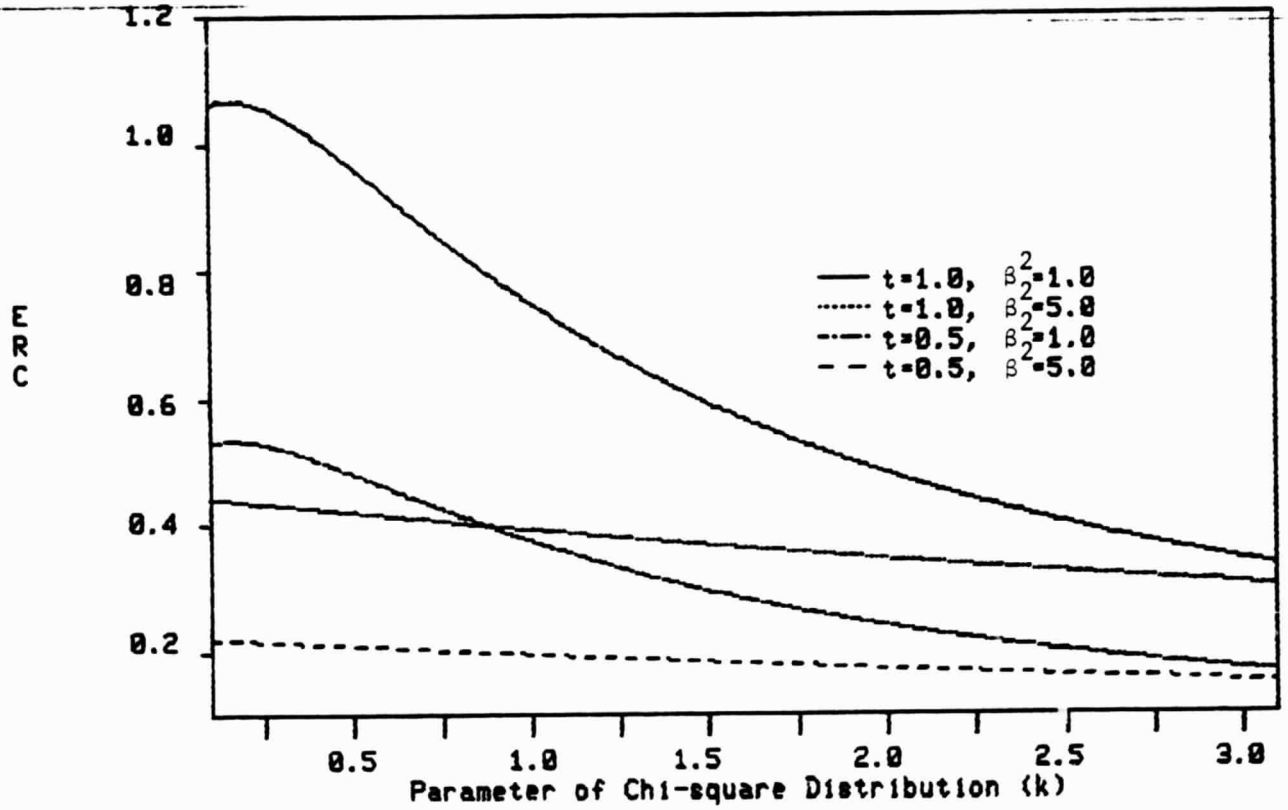


Fig 4.1: Asymptotic MSE comparisons between LS and EV estimators, Beta-squared=1.0

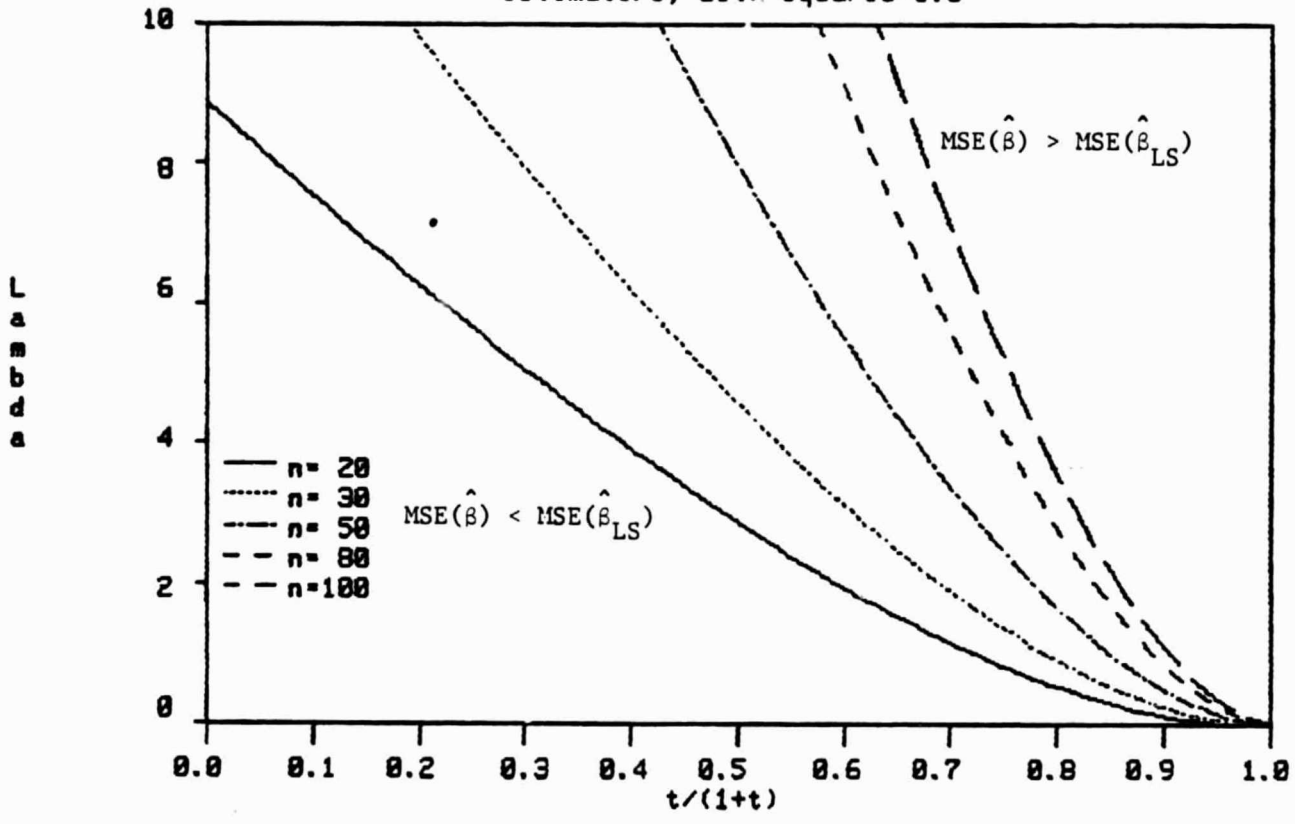


Fig 4.2: Asymptotic MSE comparisons between LS and EV estimators, Beta-squared=5.0

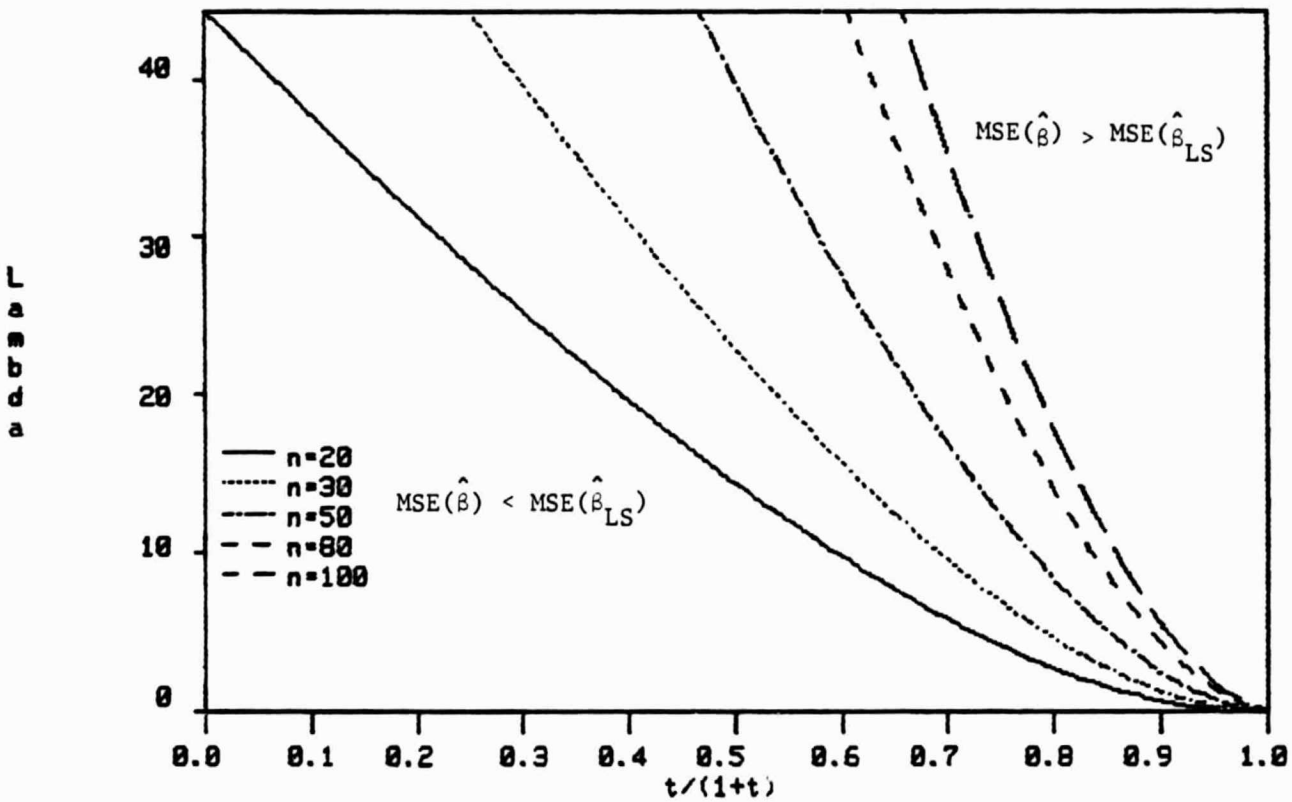


Fig 4.3: Ratio of Asymptotic MSE of EV and LS estimators

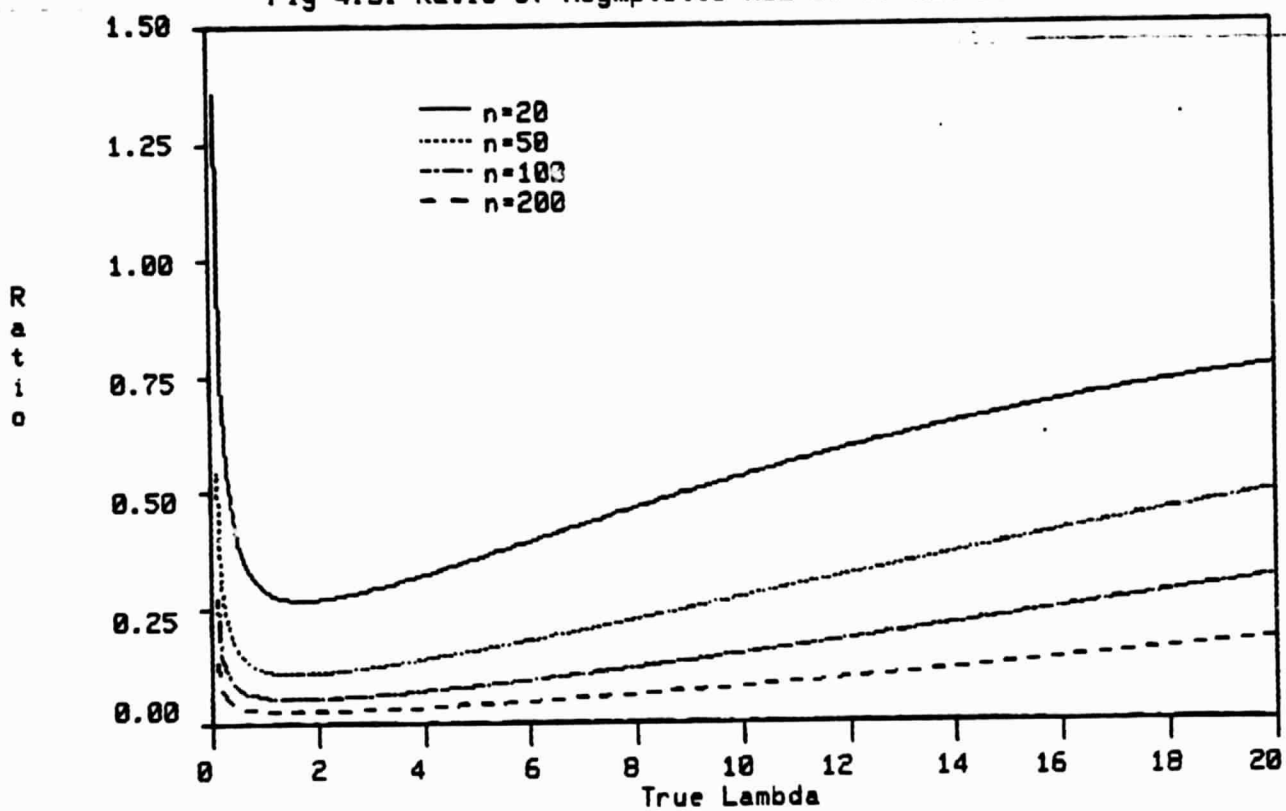


Fig 4.4: Ratio of Asymptotic MSE of EV and LS estimators
Lambda = 1.0

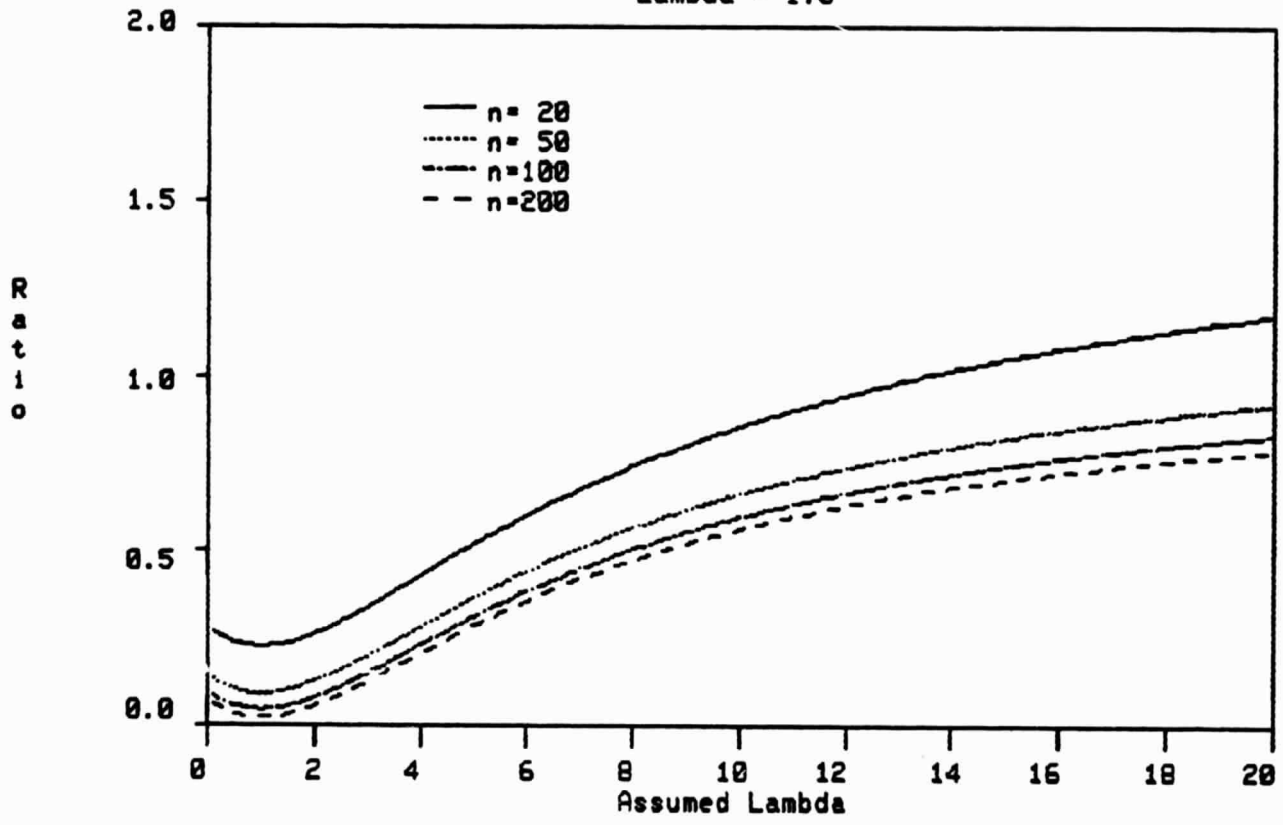


Fig 4.5: Ratio of Asymptotic MSE of EV and LS estimators
Lambda = 6.0

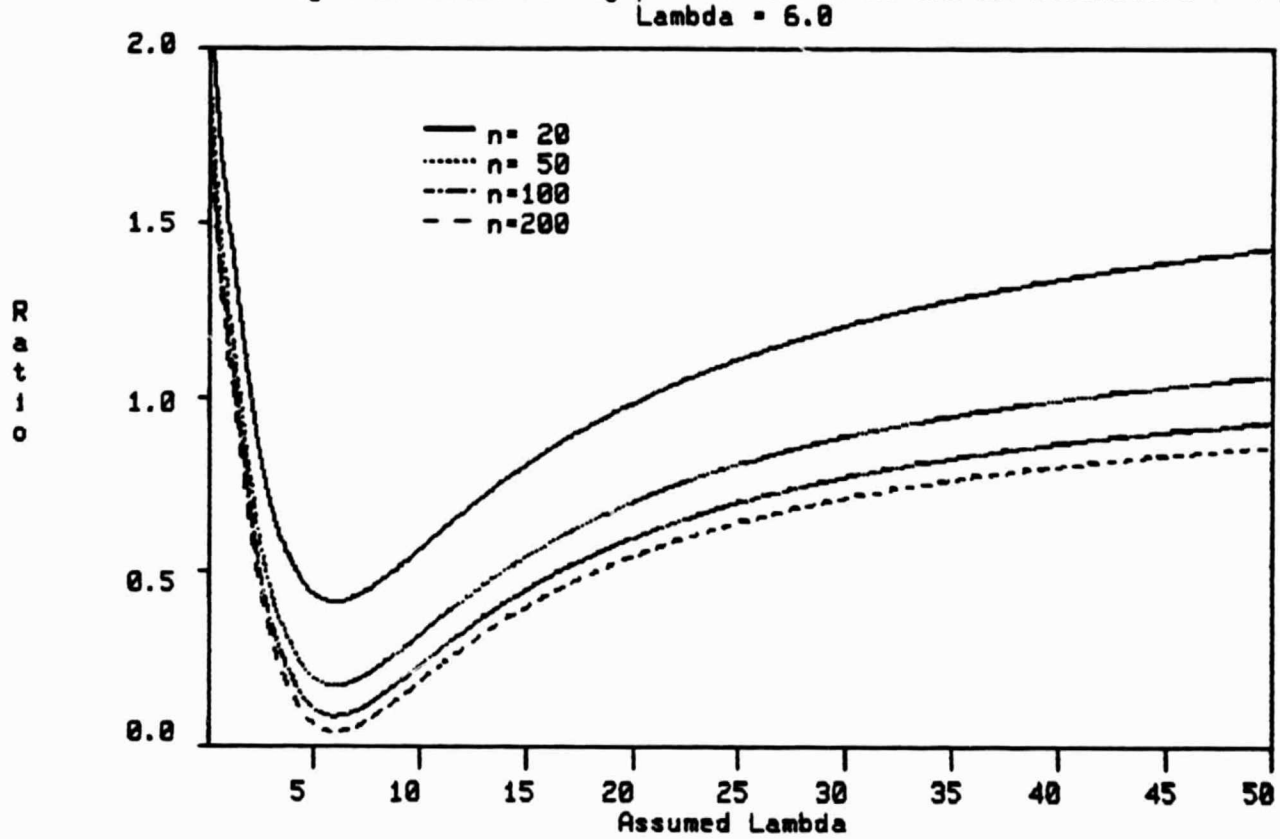
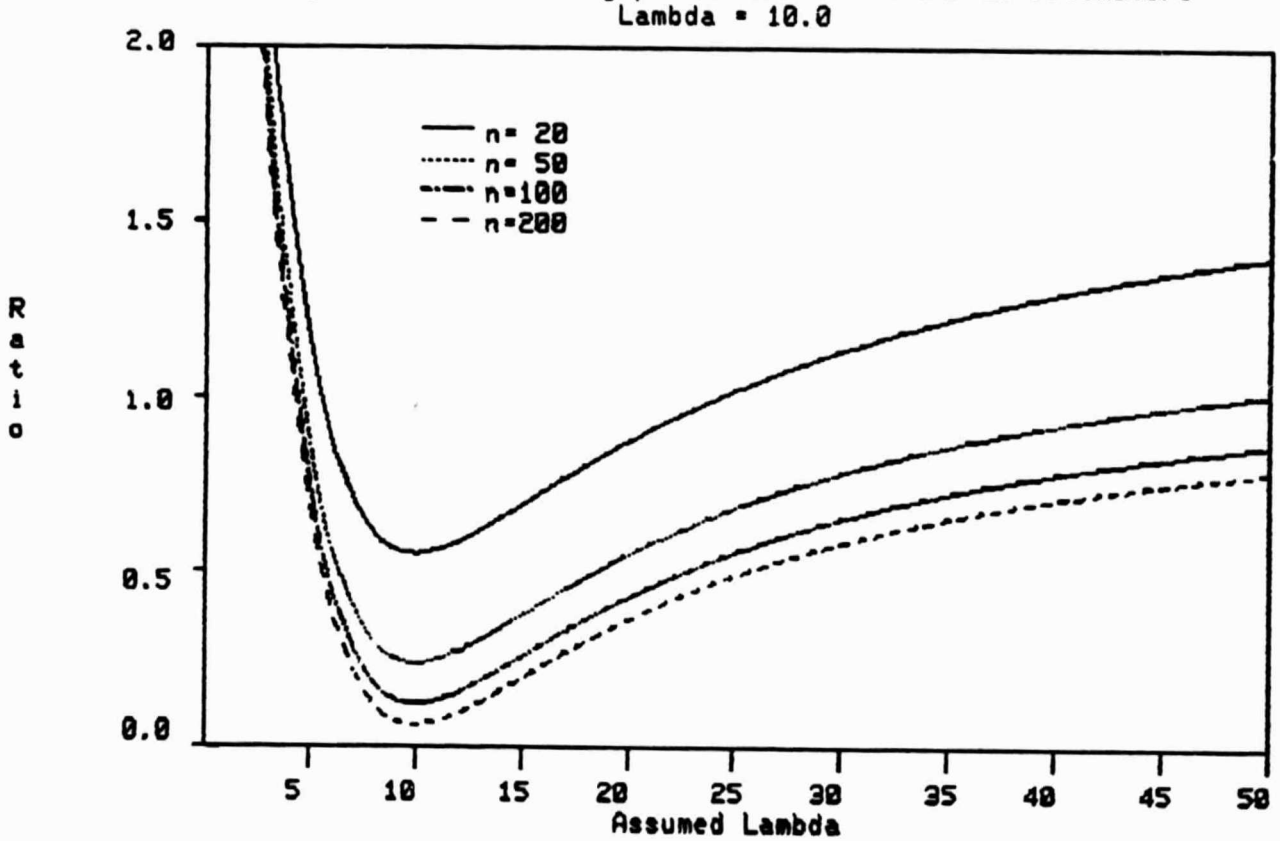


Fig 4.6: Ratio of Asymptotic MSE of EV and LS estimators
Lambda = 10.0



TOWARD A BALANCED ASSESSMENT OF COLLINEARITY DIAGNOSTICS

Richard F. Gunst*

Periodically it is wise to review the foundation upon which statistical methodology is based. With the availability of mainframe and micro computer technology there is too great a tendency to become more enamored with the sophistication with which statistical analyses can be reported than with the theoretical underpinnings of the results. Professor Belsley's article contributes to a growing number of survey papers which attempt to refocus attention on the assumptions underlying regression methodology as it is practiced today (e.g., Draper and Van Nostrand 1979; Smith and Campbell 1980; Hocking and Pendleton 1983). These articles are especially noteworthy because they force investigators to confront fundamental questions relating to one of the most difficult and controversial problems facing data analysts: redundant predictor variables in a regression analysis.

Professor Belsley criticizes the prevailing practice of centering predictor variables (usually followed by scaling to unit length) prior to assessing the presence and effects of collinearity. He clarifies the position of Belsley, Kuh and Welsch (1989) that predictor variables should be scaled to unit length but not centered prior to diagnosing collinearity. He argues unequivocally that collinearity diagnostics are only

meaningful when interpreted in terms of "basic variables" which are "structurally interpretable." In keeping with the preference endorsed in his book, he stresses the use of the condition index as the only appropriate measure of collinearity.

Without hesitation I laud Professor Belsley's effort to redress the lack of attention to the role of centering in discussions of collinearity and his effort to create a framework within which collinearity can be more rigorously examined. If I differ with him on any of the issues which he raises, my divergence of opinion rests primarily with the dogmatic insistence that there is one correct technique within which discussions of collinearity must be straightjacketed. Rather, I believe that many of the technical issues he raises are related more to one's perspective, education, and experience than necessarily to a correct technique for the proper assessment of collinearity.

1. CONFLICTING PERSPECTIVES

Although Professor Belsley repeatedly cautions against centering when diagnosing collinearity, he is careful to point out that there are legitimate circumstances under which centering is appropriate. As has been argued elsewhere (Hocking 1983, with discussion), it is common practice to center all experimental designs and attendant analyses when fitting response surfaces. Similarly, Marquardt (1980) argues that polynomial regression

coefficients derive their interpretability only when predictor variables are centered. Bradley and Srivastava (1979) stress that centered, symmetrically-located and equally-spaced values of the predictor variables should be selected in any polynomial regression analysis in which the investigator can control the values of the variates. Thus there are wide classes of regression problems for which centering is considered essential, even if the data are collinear.

The major difference between the above illustrations and the arguments posed by Professor Belsley is one of perspective. The above illustrations are most relevant in industrial settings where controlled experimentation is prevalent and constant terms are known to be necessary for adequate model fits. Observation rather than experimentation is more common in the economic studies to which Professor Belsley alludes in his example of consumption functions. In observational studies it is not necessarily assumed that constant terms are inherent to correct model specification (e.g., consumption can only be zero when the constant term is zero). Each of these perspectives should be recognized as legitimate when appropriate.

Centering can be either beneficial or detrimental regardless of whether one's perspective is derived from industrial experimentation or observational studies. Centering replaces each predictor variable with the residuals from a least squares regression of that variable on the constant term. In an intuitive sense,

centered predictor variables contain no common information with the constant term of the model. In addition, centering alters the constant term:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon \qquad (1)$$

becomes

$$y = \beta_0 + \beta_1 w_1 + \beta_2 w_2 + \ldots + \beta_p w_p + \varepsilon$$

where

$$\beta_0 = \alpha + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 + \ldots + \beta_p \bar{x}_p \quad \text{and} \quad w_j = x_j - \bar{x}_j.$$

Inferences on β_0 are all but meaningless since the $\bar{x}_j$ are data-dependent; an exception sometimes arising when the predictor-variable values are predetermined in a designed experiment. In general, then, if one wishes to make inferences on the level of the response variable (including tests for no-intercept models), centering is pointless. On the other hand, if one wishes to draw inferences on whether the predictor variables contribute to the fit of the response variable in addition to the constant term (i.e., the response variability is not simply due to random fluctuation about its level) then centered predictor variables are essential.

In this latter setting one need not sacrifice diagnostic information about possible collinearity of the predictor variables with the constant term. The estimated standard error of the constant term of model (1) can be expressed as

$$\begin{aligned} s.\hat{e}.(\hat{\beta}_0) &= \hat{\sigma}(n - \underline{1}'X(X'X)^{-1}X'\underline{1})^{-1/2} \\ &= (\hat{\sigma}/n^{1/2})(1-R_0^2)^{-1/2} \end{aligned}$$

where R_0^2 is the coefficient of determination when the constant term is regressed on the other predictor variables. Note that if in the definition of the model the columns of X are centered, as can occur in an experimental design, $s.\hat{e}.(\hat{\beta}_0) = \hat{\sigma}/n^{1/2}$; otherwise, $s.\hat{e}.(\hat{\beta}_0) > \hat{\sigma}/n^{1/2}$. Consequently,

$$s.\hat{e}.(\hat{\beta}_0)^2/(\hat{\sigma}^2/n) = 1/(1-R_0^2)^{-1}$$

is the variance inflation factor for the constant term of the model. For the example in Section 1,

$$s.\hat{e}.(\hat{\beta}_0)^2/(\hat{\sigma}^2/n) = (.784)^2/([.0055]^2/20) > 400{,}000.$$

There is clearly a collinearity problem among the predictor variables and the constant term. Centering does not demean the collinearity diagnostics, one must simply understand the nature of centering and know where to look for the appropriate diagnostic.

2. MEASURING COLLINEARITY

Comparisons of collinearity measures provide valuable guidelines for data analysts. Farrar and Glauber (1967), Leamer (1973), Mason, Gunst, and Webster (1975), Willan and Watts (1978), and Belsley, Kuh, and Welsch (1980) describe a wide variety of important collinearity diagnostics; however, rarely will any single collinearity measure completely characterize the nature and effects of collinear predictor variables. Some measures are appropriate for assessing the sensitivity of least squares estimates to minor perturbations of the input data (condition indices), others more readily measure

the effects of collinearity on the variances of the estimators (variance inflation factors), still others aid in identifying the nature of the collinearities (predictor-variable correlations, eigenvalues and eigenvectors of suitably scaled, perhaps centered, X'X matrices). Table 1 lists selected collinearity measures from the above references according to possible usage.

[Insert Table 1]

Just as there is no monopoly by any single collinearity measure on usefulness, collinearity itself is difficult to define. Professor Belsley stresses conditioning as a descriptor of collinearity. This author's preference is to define collinearity by analogy with the algebraic definition of linear dependence among a (normalized) set of vectors (Gunst 1983):

Defn. A collinearity is said to exist among the columns of $X = [\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_p]$ if for a suitably small predetermined $\eta > 0$ there exists constants $c_1, c_2, \ldots, c_p$, not all of which are zero, such that

$$c_1\underline{x}_1 + c_2\underline{x}_2 + \ldots + c_p\underline{x}_p = \underline{\delta} \text{ with } ||\underline{\delta}|| < \eta \cdot ||\underline{c}||.$$

Neither this definition nor any other which can be offered is entirely satisfactory (e.g., How small should η be? How large should a condition index be?) but each is a meaningful concept to many data analysts, depending again on background and experience. Hocking and Pendleton's (1983) "picket fence" analogy is a marvelously simple geometric explanation of collinearity which can be more useful than either of the above technical

definitions when one must characterize predictor-variable redundancies to those who have limited statistical training.

Even from a strictly analytic point of view there are difficulties with all measures of collinearity, difficulties which limit the global utility of each diagnostic. Belsley, Kuh, and Welsch contend that small eigenvalues of X'X are inadequate measures of collinearity since perfectly-conditioned matrices of the form

$$X'X = \begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix}$$

can have arbitrarily small eigenvalues for small values of $\alpha > 0$ even though the two columns of X are orthogonal (no collinearity). Condition indices suffer from much the same problem if one examines perfectly-conditioned matrices of the form

$$X'X = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}$$

and $\beta > 0$ is allowed to become arbitrarily small while $\alpha > 0$ is held fixed. Note that if one scales the columns of X to unit length both of the above matrices are identity matrices for all $\alpha, \beta > 0$ and both the eigenvalues of X'X and the condition index of X will correctly diagnose the perfect conditioning.

3. STRUCTURAL INTERPRETABILITY

A major contribution of this article, one with whith I am in wholehearted agreement, is its focus on correct model formulation.

Model formulation receives scant attention in regression textbooks, most of the emphasis being directed to variable transformations. Structural interpretability demands that careful consideration be directed toward the initial specification of each variable in a regression model. Professor Belsley specifically directs his comments to centering but the overall admonition which he conveys is more general.

The question of variable definition is a difficult one. It is too frequent that one encounters "proxy" or "surrogate" variables used in place of the real quantities one seeks but is unable to measure. The rationalization that this is the best one can do leads to acceptance of arbitrary variable definitions in many other circumstances. For example, few worry about whether temperature is expressed in degrees Celsius or Fahrenheight. Yet this is perhaps the traditional example one uses to distinguish interval- from ratio-scaled variables in introductory statistics courses, a univariate version of structural interpretability.

Structural interpretability as a universally-accepted principle in model formulation must await refinement. The relevance of structural interpretability, at least with regard to centering, is not clear when quantities such as "beta-weights" are the intended goal of a regression analysis. Calculation of these quantities requires that predictor variables be standardized, a transformation which destroys structural interpretability. Extension to polynomial models, models which include interaction terms, and nonlinear models--

admittedly topics which are beyond the scope of the present manuscript -- are additional important refinements which await clarification. Nevertheless, the warning is clear: challenges to interpretability are inevitable when model formulation, including centering, is slighted.

Finally, structural interpretability is a concept which is meaningful without any reference to collinearity. The application of structural interpretability to regression implies that a constant term is included in the model if appropriate, not of necessity. Collinearity is a separate issue which becomes relevant only after proper model formulation. In order to properly diagnose collinearity variates should be centered or standardized as required to apply appropriate analytical techniques.

4. THE EXAMPLE

Smith and Campbell (1980) discuss clearly the relationship between (linear) predictor-variable transformations and the disguising of collinearities as small predictor-variable variances. Mullett (1976) demonstrates that the ill-effects usually associated with collinearities can be produced by other causes, including small predictor-variable variances. These discussions have particular relevance to collinearities involving the constant term.

Earlier the collinearity among the three predictor variables in the example was shown to be diagnosable from the standard error of the estimated coefficient of the constant term. That the "non-

constant" predictor variables are essentially constant is apparent from their coefficients of variation ($s_j/\bar{x}_j$): each is approximately .0023. Regardless of whether each is correlated with the other, a coefficient of variation this small calls for immediate investigation of collinearity if one's intent is to evaluate each of the predictor variables for its predictive ability without regard to the presence or absence of the others, including a constant term.

A collinearity with the constant term occurs either because two or more of the nonconstant predictors are (reasonably) variable and some linear combination of them is essentially constant or because individual variates are essentially constant. The former situation can be detected from an analysis of the centered (standardized) variates, the latter from the coefficients of variation (often just from the standard deviations). In either case, if collinearity with the constant term is a concern, examination of the standard error of the constant term will readily reveal the existence of a problem.

4. FINAL REMARKS

This manuscript is an excellent example of the dialogue which should periodically review the foundations upon which regression methodology is based. While differences of opinion will inevitably arise, separation of the fundamental issues from personal preference is important. I am in fundamental agreement with Professor Belsley

on what I perceive to be the key issues in his article: (i) model formulation, using concepts like structural interpretability, is essential for meaningful inferences from a regression analysis, (ii) careful consideration should be given to whether collinearity with the constant term is important to detect, and (iii) if so, collinearity diagnostics which enable such detection must be examined. While we may prefer alternative collinearity diagnostics, we seek the same goal with equally-effective diagnostic techniques.

REFERENCES

BELSLEY, D.A., KUH, E. and WELSCH, R. E. (1980), *Regression Diagnostics*, New York: John Wiley and Sons, Inc.

BRADLEY, R. A. and SRIVASTAVA, S. S. (1979), "Correlation in Polynomial Regression," *The American Statistician*, 33, 11-14.

DRAPER, N. R. and VAN NOSTRAND, R. C. (1979), "Ridge Regression and James-Stein Estimation: Review and Comments," *Technometrics*, 21, 451-466.

FARRAR, D. E. and GLAUBER, R. R. (1967), "Multicollinearity in Regression Analysis: The Problem Revisited," *Review of Economics and Statistics*, 49, 92-107.

GUNST, R. F. (1983), "Regression Analysis with Multicollinear Predictor Variables: Definition, Detection and Effects," *Communications in Statistics, A12* (Special Issue on Statistical Reviews), 2217-2260.

HOCKING, R. R. (1983), "Developments in Linear Regression Methodology: 1959-1982," *Technometrics, 25*, 219-249.

HOCKING, R. R. and PENDELTON, O. J. (1983), "The Regression Dilemma," *Communications in Statistics, A12*, 497-527.

LEAMER, E. E. (1973), "Multicollinearity: A Bayesian Interpretation," *The Review of Economics and Statistics, 55*, 371-380.

MARQUARDT, D. W. (1980), "You Should Standardize the Predictor Variables in Your Regression Models," *American Statistician, 75*, 87-91.

MASON, R. L., GUNST, R. F. and WEBSTER, J. T. (1975), "Regression Analysis and Problems of Multicollinearity," *Communications in Statistics, 4*, 277-292.

SMITH, G. and CAMPBELL, F. (1980), "A Critique of Some Ridge Regression Methods," *Journal of the American Statistical Association, 75*, 74-81.

WILLIAN, A. R. and WATTS, D. G. (1978), "Meaningful Multicollinearity Measures," *Technometrics, 20*, 407-412.

AUTHOR'S FOOTNOTE

*Richard F. Gunst is Chairman, Department of Statistics, Southern Methodist University, Dallas, TX. 75275. This research was supported in part by the National Aeronautics and Space Administration under Contract No. NCC 9-9.

TABLE 1. Selected Collinearity Measures

Detection Measures	Estimator Effects	Precision
Predictor-Variable Correlations	Condition Indices	Variance Inflation Factors
Variance Inflation Factors	Estimator Correlations	s.e.($\hat{\beta}_j$), s.e.($\hat{y}$)
Eigenvalues, Eigenvectors of X'X	Curve Decolletages	Variance Decomposition Proportions
Condition Indices		Volumes of Confidence Ellipsoids

REGRESSION DIAGNOSTICS AND APPROXIMATE INFERENCE PROCEDURES FOR PENALIZED LEAST SQUARES ESTIMATORS

R. L. Eubank and R. F. Gunst*

ABSTRACT

Generalizations of least squares diagnostic techniques are presented for a class of penalized least squares estimators. Efficient computation of these diagnostics is afforded by expressions which relate coefficient estimates and residuals from fits to subsets of the data to the corresponding quantities from a fit to the complete data set. From these expressions approximate confidence intervals and test statistics can be obtained using jackknife and bootstrap procedures. Applications are discussed for the special cases of smoothing splines and ridge regression.

KEY WORDS

Bootstrap confidence intervals; Jackknife confidence intervals; Leverage values; Ridge regression; Smoothing splines; Studentized residuals.

AUTHOR'S FOOTNOTE

* R. L. Eubank is Assistant Professor and R. F. Gunst is Associate Professor and Chairman, Department of Statistics, Southern Methodist University. This research was supported in part by Office of Naval Research Contract No. N00014-82-K-0207 and by the National Aeronautics and Space Administration under Contract No. NCC 9-9. The authors would like to thank Mr. Tom Carmody for furnishing Figure 1 and Dr. Paul Speckman for helpful discussions on the material in this manuscript.

REGRESSION DIAGNOSTICS AND APPROXIMATE INFERENCE PROCEDURES FOR PENALIZED LEAST SQUARES ESTIMATORS

R. L. Eubank and R. F. Gunst

1. INTRODUCTION

Regression diagnostics are an integral component of comprehensive regression modeling efforts, in large part because of recent theoretical advances which lead to computational efficiency. With few exceptions (a notable one being Pregibon (1981)) these advances have been restricted to ordinary least squares (OLS) estimation for linear models. In this paper diagnostic techniques are extended to a class of penalized least squares estimators which include smoothing splines and ridge regression estimators as special cases. An additional benefit of these results is the ability to efficiently compute jackknife confidence intervals and other inferential statistics for model parameters.

Let $y = (y_1,\ldots,y_n)'$ be a vector of observed responses which follow the model

$$y = \eta + \varepsilon \,, \tag{1.1}$$

where $\eta = (\eta_1,\ldots,\eta_n)'$ is a vector of unknown constants and $\varepsilon = (\varepsilon_1,\ldots,\varepsilon_n)'$ is a vector of zero mean, uncorrelated errors with

common variance σ^2. It is assumed that η is to be approximated by a linear form $X\beta$ where X is a known nxp matrix of rank $p \leq n$ having ith row x_i' and $\beta = (\beta_1,\ldots,\beta_p)'$ is a vector of parameters wnich is to be estimated. The class of estimators which are investigated in this article are those obtained as the solution to

$$\min_{\beta}\{\Sigma_{j=1}^{n}(y_j - x_j'\beta)^2 + \lambda\beta'Q\beta\}, \quad \lambda \geq 0 , \tag{1.2}$$

with Q denoting an arbitrary positive (semi-) definite matrix. For a given Q, X, and λ, expression (1.2) has a unique solution:

$$\tilde{\beta} = C(\lambda)y , \tag{1.3}$$

where

$$C(\lambda) = (X'X + \lambda Q)^{-1}X' . \tag{1.4}$$

The estimator $\tilde{\beta}$ is termed a penalized least squares estimator of β. Observe that when $\lambda = 0$, $\tilde{\beta}$ reduces to the OLS estimator

$$\hat{\beta} = (X'X)^{-1}X'y .$$

At the other extreme, if Q is positive definite $\tilde{\beta} \to \underline{0}$ as $\lambda \to \infty$. In many instances it is preferable to use a value of λ between these two extremes and a variety of methods are available for estimating its value from data. For example, Golub, Heath and Wahba (1979) discuss generalized cross-validation (GCV) as well as other data-driven methods for selecting λ.

It is often reasonable to make the stronger assumption that $\eta = X\beta$ under which model (1.1) becomes the linear regression model

$$y = X\beta + \varepsilon . \tag{1.6}$$

When this model holds and no further assumptions are made, $\tilde{\beta}$ will be termed a generalized ridge regression estimator of β; however,

the results presented below are of sufficient generality to include cases in which the η_j represent values from an unknown regression function, η, which is to be estimated nonparametrically. When appropriately formulated (see Section 5) the smoothing spline estimator of η is seen to be a special case of estimator (1.3).

As with ordinary least squares, the penalized least squares "hat matrix" (see Hoaglin and Welsch 1978) provides important diagnostic information about the influence of individual observations (y_i, x_i') on the associated prediction equation. The hat matrix corresponding to $\tilde{\beta}$ is defined to be

$$H(\lambda) = \{h_{ij}(\lambda)\} = XC(\lambda) \ . \qquad (1.7)$$

This matrix transforms the response vector y to the vector of fitted values, $\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_n)'$; i.e.,

$$\tilde{y} = H(\lambda)y \ .$$

The element $h_{ij}(\lambda)$ is a direct measure of the influence of y_j on the fit to y_i. In particular, the "leverage value" $h_{ii}(\lambda)$ measures the influence of y_i on its own prediction.

This study of the estimator class (1.3) begins with a derivation of some of the properties of $H(\lambda)$ in Section 2. In Section 3 techniques are presented for computing estimates and fitted values when observations are deleted from the data set. The results of this section are applied, in Section 4, to obtain approximate inference procedures for the parameter vector β and to derive diagnostic measures for detecting influential observations. Specific applications to nonparametric estimation by smoothing splines and to ridge regression estimators are detailed in Section 5. Con-

cluding remarks are made in Section 6.

2. LEVERAGE VALUES FOR PENALIZED LEAST SQUARES

In this section certain properties of the hat matrix $H(\lambda)$ will be derived. It will be seen that the characteristics of its elements are closely related to those of the hat matrix H for the corresponding OLS estimator:

$$H = \{h_{ij}\} = X(X'X)^{-1}X' . \tag{2.1}$$

Since H in equation (2.1) is a (orthogonal) projection operator, the following properties are easily proven:

$$\begin{aligned} &\text{i)} \quad 0 \leq h_{ii} \leq 1 \\ &\text{ii)} \quad -1 \leq h_{ij} \leq 1, \quad i \neq j \\ &\text{iii)} \quad h_{ij} = 1 \implies h_{ij} = 0, \quad i \neq j . \end{aligned} \tag{2.2}$$

When X contains a constant column, somewhat sharper results are provided by

$$\begin{aligned} &\text{i)}' \quad n^{-1} \leq h_{ii} \leq 1 \\ &\text{ii)}' \quad -(n-1)n^{-1} \leq h_{ij} \leq 1, \quad i \neq j \\ &\text{iii)}' \quad h_{ii} = 1 \iff h_{ij} = 0 , \quad i \neq j . \end{aligned} \tag{2.3}$$

Extreme rows of X result in large leverage values. The rough cutoff of $h_{ii} > 2p/n$ suggested by Hoaglin and Welsch (1978) is often used to identify such rows. Note from iii) and iii)' that, as $h_{ii} \to 1$, $h_{ij} \to 0$, $i \neq j$ and $\hat{y}_i = x_i'\hat{\beta} \to y_i$, indicating that an observation with a large leverage value will tend to dominate its own fit.

For $\lambda > 0$, $H(\lambda)$ is no longer a projection matrix. The following

theorem establishes bounds for the elements of $H(\lambda)$ as a function of the corresponding elements of H, thereby providing an analog of properties i) and ii) in equation (2.2).

Theorem 2.1. The elements of $H(\lambda)$ satisfy

$$|h_{ij}(\lambda)| \leq (1 + \lambda d_1)^{-1}\{h_{ii}h_{jj}\}^{1/2} \qquad (2.4)$$

where d_1 is the smallest eigenvalue of $(X'X)^{-1}Q$.

Proof. Using the spectral decomposition (eg. Kshirsagar 1972, Chapter 7) of X write $X = UL^{1/2}Z'$, where $L = \text{diag}(\ell_1,\ldots,\ell_p)$ is a diagonal matrix containing the nonzero eigenvalues of XX' (and X'X) in ascending order, and $U = [u_1,\ldots,u_p]$ and Z are the corresponding matrices of eigenvectors of XX' and X'X, respectively. $H(\lambda)$ can now be expressed as

$$H(\lambda) = U(I + \lambda L^{-1/2}Z'QZL^{-1/2})^{-1}U' \; . \qquad (2.6)$$

Let $0 \leq d_1 \leq d_2 \leq \ldots \leq d_p$ denote the eigenvalues of $L^{-1/2}Z'QZL^{-1/2}$ (which are also the eigenvalues of $(X'X)^{-1}Q$). Using $\Gamma = [\gamma_1,\ldots,\gamma_p]$ to denote the corresponding matrix of eigenvectors, individual elements of $H(\lambda)$ can now be represented as

$$h_{ij}(\lambda) = \Sigma_{r=1}^{p} b_{ir}b_{jr}(1+\lambda d_r)^{-1}, \quad b_{kr} = u_k'\gamma_r. \qquad (2.7)$$

Application of the Cauchy-Schwarz inequality in equation (2.7) along with the ordering of the d_r completes the proof. □

Theorem 2.1 and its proof have several important consequences. First, it furnishes tighter bounds for the elements of $H(\lambda)$ than the inequalities in equation (2.2); i.e.,

$$\begin{aligned} &\text{i)} \quad 0 \leq h_{ii}(\lambda) \leq (1 + \lambda d_1)^{-1} \\ &\text{ii)} \quad -(1 + \lambda d_1)^{-1} \leq h_{ij}(\lambda) \leq (1 + \lambda d_1)^{-1}, \; i \neq j \; . \end{aligned} \qquad (2.8)$$

In addition, from equation (2.7), it is apparent that $h_{ii}(\lambda)$ is monotonically decreasing with λ from $h_{ii}(0) = h_{ii}$ to $h_{ii}(\infty)$. Note that in general $h_{ii}(\infty) > 0$ unless $d_1 > 0$; when $d_1 > 0$, $h_{ii}(\infty) = 0$. Since $h_{ij}(\lambda)$ is continuous in λ, standard results from calculus can be used to show that for λ sufficiently small (large) $h_{ij}(\lambda)$ will have the same sign as h_{ij} ($h_{ij}(\infty)$) provided that $h_{ij} \neq 0$ ($h_{ij}(\infty) \neq 0$).

Two important special cases occur when (i) $0 = d_1 = \ldots = d_m < d_{m+1} \leq \ldots \leq d_p$ and (ii) $Q = I$. These special cases have applications to smoothing splines and ridge regression, respectively, which will be explored in Section 5. The important details are summarized in the following two corollaries.

Corollary 1. Suppose $0 = d_1 = \ldots = d_m < d_{m+1} \leq \ldots \leq d_p$ and define $h_{ij}(\infty) = \Sigma_{r=1}^{m} b_{ir} b_{jr}$, where the b_{kr} are as in equation (2.7). Then

$$h_{ii}(\infty) + (1+\lambda d_p)^{-1} \sum_{r=m+1}^{p} b_{ir}^2 \leq h_{ii}(\lambda) \leq h_{ii}(\infty) + (1+\lambda d_{m+1})^{-1} \sum_{r=m+1}^{p} b_{ir}^2. \tag{2.9}$$

Corollary 2. If $\tilde{\beta} = (X'X + \lambda I)^{-1} X'y$ then

$$|h_{ij}(\lambda)| \leq \ell_p (\ell_p + \lambda)^{-1} \{h_{ii} h_{jj}\}^{1/2}$$

where ℓ_p is the largest eigenvalue of $X'X$. The upper bound for the ith leverage value, viz. $\ell_p (\ell_p + \lambda)^{-1}$, is obtained when $x_j' = \ell_p^{1/2} z_p'$ where z_p is the eigenvector corresponding to ℓ_p.

3. DELETING OBSERVATIONS FROM AN ESTIMATOR

The development of exact tests and interval estimates for β using the penalized least squares estimator $\tilde{\beta}$ is a difficult, and

as yet unresolved, problem. In contrast, approximate techniques based on nonparametric procedures such as the jackknife and bootstrap are easy to propose but their practicality depends on the ability to efficiently perform the necessary calculations. In this section a simple method of deleting observations from $\tilde{\beta}$ is derived which requires no refitting of the data. This is found, in Section 4, to make the use of inference techniques such as jackknife confidence regions for $\tilde{\beta}$ a practical alternative and to allow a generalization of several types of regression diagnostic measures to the penalized least squares setting.

For $q \leq n-p$ let $J = \{j_1,\ldots,j_q\}$ be a subset of the indices $\{1,\ldots,n\}$ and let $\tilde{\beta}^{(J)}$ represent the coefficient estimates obtained using only those (y_j, x_j') with $j \notin J$. The following theorem provides a partial characterization of $\tilde{\beta}^{(J)}$.

Theorem 3.1. Let $\tilde{\beta}^{[J]}(w_{j_1},\ldots,w_{j_q})$ solve

$$\min_{\beta}\{\sum_{j\notin J}(y_j-x_j'\beta)^2 + \sum_{j\in J}(w_j-x_j'\beta)^2 + \lambda\beta'Q\beta\} \tag{3.1}$$

and define $\tilde{y}_i^{(J)} = x_i'\tilde{\beta}^{(J)}$, $i = 1,\ldots,n$. Then,

$$\tilde{\beta}^{[J]}(\tilde{y}_{j_1}^{(J)},\ldots,\tilde{y}_{j_q}^{(J)}) = \tilde{\beta}^{(J)} . \tag{3.2}$$

Theorem 3.1 has the consequence that $\tilde{\beta}^{(J)}$ can be obtained by applying $C(\lambda)$ to a "new data vector" wherein y_j has been replaced by $\tilde{y}_j^{(J)}$ for all $j\in J$. This would seem to presuppose knowledge of $\tilde{\beta}^{(J)}$; however, such is not the case and in many cases of interest it is possible to compute the $\tilde{y}_j^{(J)}$ without explicit computation of $\tilde{\beta}^{(J)}$. This property follows by application of the next theorem.

Theorem 3.2. The values $\tilde{y}_j^{(J)}$, $j\in J$, satisfy the linear equation system

$$\tilde{y}_i^{(J)} - \Sigma_{j\in J} h_{ij}(\lambda)\tilde{y}_j^{(J)} = \tilde{y}_i - \Sigma_{j\in J} h_{ij}(\lambda) y_j$$

$$= \Sigma_{j\notin J} h_{ij}(\lambda) y_j \ , \quad i \in J \ . \qquad (3.3)$$

Proof of Theorems 3.1-3.2. Set $w_j = \tilde{y}_j^{(J)}$. Proof of Theorem 3.1 is provided by the following inequalities:

$$\Sigma_{j\notin J}(y_j - x_j'\tilde{\beta}^{(J)})^2 + \Sigma_{j\in J}(w_j - x_j'\tilde{\beta}^{(J)})^2 + \lambda\tilde{\beta}^{(J)\prime} Q\tilde{\beta}^{(J)}$$

$$= \Sigma_{j\notin J}(y_j - x_j'\tilde{\beta}^{(J)})^2 + \lambda\tilde{\beta}^{(J)\prime} Q\tilde{\beta}^{(J)} \leq \Sigma_{j\notin J}(y_j - x_j'\beta)^2 + \lambda\beta' Q\beta$$

$$\leq \Sigma_{j\notin J}(y_j - x_j'\beta)^2 + \Sigma_{j\in J}(w_j - x_j'\beta)^2 + \lambda\beta' Q\beta \ .$$

To verify equation (3.3) note that $x_i'\tilde{\beta}^{[J]}(w_{j_1},\ldots,w_{j_q})$ is linear in w_j, $j\in J$, and can, therefore, be written as

$$x_i'\tilde{\beta}^{[J]}(w_{j_1},\ldots,w_{j_q}) = x_i'\tilde{\beta} + \Sigma_{j\in J} h_{ij}(\lambda)(w_j - y_j) \ . \qquad (3.4)$$

Letting $w_j = x_j'\tilde{\beta}^{(J)}$ gives the desired result. □

To illustrate uses for Theorem 3.1-3.2 confine attention, for the moment, to the instance $q = 1$, $J = \{j\}$ for some $j \in \{1,\ldots,n\}$. To distinguish this important special case the notation

$$\tilde{\beta}^{[j]} = \tilde{\beta}^{(J)} \qquad (3.5)$$

and

$$\tilde{y}_i^{[j]} = x_i'\tilde{\beta}^{[j]} \qquad (3.6)$$

is utilized. Application of Theorem 3.2 to this special case yields the following expression for $\tilde{y}_j^{[j]}$:

$$\tilde{y}_j^{[j]} = (\tilde{y}_j - h_{jj}(\lambda)y_j)/(1-h_{jj}(\lambda)). \qquad (3.7)$$

This relationship explicitly demonstrates the ability to obtain each of the $\tilde{y}_j^{[j]}$ without refitting the model.

The term "deleted residual" will be used to designate the difference $y_j - \tilde{y}_j^{[j]}$. Equation (3.7) provides an efficient computational form for the deleted residual; viz.,

$$e_{[j]} \equiv y_j - \tilde{y}_j^{[j]} = e_j/(1-h_{jj}(\lambda)), \quad j = 1,\ldots,n, \quad (3.8)$$

where e_j is the jth residual from the fit to the entire data set:

$$e_j = y_j - \tilde{y}_j, \quad j = 1,\ldots,n. \quad (3.9)$$

Substituting equation (3.8) into equation (3.2) yields

$$\tilde{\beta}^{[j]} = \tilde{\beta} - c_j(\lambda)e_{[j]}, \quad j = 1,\ldots,n, \quad (3.10)$$

where $c_j(\lambda)$ is the jth column of $C(\lambda)$.

Formulas (3.8) and (3.10) include as special cases the equivalent expressions for ordinary least squares, $\lambda = 0$ (e.g., Beckman and Trussel 1974; Hoaglin and Welsch 1978). In the case of smoothing splines equation (3.8) was established by Craven and Wahba (1979) using a method of proof similar to the one employed here.

4. INFERENCE AND DIAGNOSTICS

Equation (3.8) provides a fundamental expression for the derivation of approximate confidence intervals to complement the point estimator $\tilde{\beta}$. Define the jth vector of pseudo-values by

$$\tilde{b}_j = n\tilde{\beta} - (n-1)\tilde{\beta}^{[j]}$$

$$= \tilde{\beta} + (n-1)c_j(\lambda)e_{[j]} . \quad (4.1)$$

Then the jackknife estimator of β based on $\tilde{\beta}$ is $\tilde{b} = n^{-1}\sum_{j=1}^{n}\tilde{b}_j$

and the variance-covariance matrix of $\tilde{\beta}$ or $\tilde{b}$ can be estimated by

$$\tilde{V} = \Sigma_{j=1}^{n}(\tilde{b}_j - \bar{\tilde{b}})(\tilde{b}_j - \bar{\tilde{b}})'/n(n-1) \ . \qquad (4.2)$$

For a linear functional $a'\beta$, an approximate $100(1-\alpha)\%$ confidence interval is provided by

$$a'\tilde{\beta} \pm Z_{\alpha/2}(a'\tilde{V}a)^{1/2} \quad \text{or} \quad a'\tilde{b} \pm Z_{\alpha/2}(a'\tilde{V}a)^{1/2} \qquad (4.3)$$

where $Z_{\alpha/2}$ is the $100(1-\alpha/2)$ percentage point of the standard normal distribution (critical values from a Student's t distribution with $n-1$ degrees of freedom could be used in place of $Z_{\alpha/2}$ in expression (4.3)). Notice that the interval estimates (4.3) can be computed using information available entirely from the original fit. When $\lambda = 0$, equations (4.1)-(4.2) reduce to formulae given in Miller (1974), Hinkley (1977a), and Efron (1982, Chapter 3) for jackknifing $\hat{\beta}$.

Diagnostic measures which parallel those utilized for ordinary least squares can also be derived as a result of (3.8) and (3.10). To do so first note that a natural estimator of σ^2 associated with the penalized least squares estimator $\tilde{\beta}$ is

$$\tilde{\sigma}^2 = \Sigma_{i=1}^{n} e_i^2/\mathrm{tr}(I-H(\lambda)) \qquad (4.4)$$

where tr denotes the matrix trace. This estimator reduces to the usual estimator of σ^2 associated with $\hat{\beta}$, namely $\hat{\sigma}^2 = \Sigma_{i=1}^{n} e_i^2/(n-p)$, when $\lambda = 0$. The estimator (4.4) has been found to be quite effective for spline smoothing by Wahba (1983). Studentized (deleted) residuals can then be defined as

$$t_{[j]} = e_j/\tilde{\sigma}_{[j]}(1-h_{jj}(\lambda))^{1/2} \qquad (4.5)$$

where $\tilde{\sigma}^2_{[j]}$ is the estimator (4.4) computed from the reduced data set

wherein (y_j, x_j') has been excluded. An explicit formula for $\tilde{\sigma}^2_{[j]}$ is

$$\tilde{\sigma}^2_{[j]} = \sum_{\substack{i=1 \\ i\neq j}}^{n} (e_i + h_{ij}(\lambda)e_{[j]})^2/\mathrm{tr}(I-H^{[j]}(\lambda)) \qquad (4.6)$$

with

$$\mathrm{tr}(I-H^{[j]}(\lambda)) = n-1 - \sum_{\substack{i=1 \\ i\neq j}}^{n} [h_{ii}(\lambda) + h_{ij}(\lambda)^2/(1-h_{jj}(\lambda))]. \qquad (4.7)$$

To prove formulas (4.6)-(4.7) observe that $\tilde{y}_i^{[j]}$ can be written as $\Sigma_{r\neq j} a_{ir} y_r$. The coefficients a_{ir} can be deduced from equation (3.2) and used to establish equation (4.7). The form of the numerator follows easily from expression (3.10).

The studentized residuals along with formulas (4.6)-(4.7) are generalizations of relations which hold when $\lambda = 0$ (e.g., Gunst and Mason 1980, Chapter 7). These residuals provide a scaled measure of how the fit to y_j changes when its value is not used to estimate β. They can, therefore, be used to detect overly influential data values. The value of $t_{[j]}$ might be compared to values from a Student's t distribution with approximately $\mathrm{tr}(I-H^{[j]}(\lambda))$ degrees of freedom. Simulation results discussed in Section 5 indicate that Student's t critical values provide a reasonably good approximation for 5% cutoff values for the $t_{[j]}$. Through similar considerations a variety of other diagnostic measures can also be suggested. One such example is

$$\mathrm{DFFITS}_j = (\tilde{y}_j - x_j'\tilde{\beta}^{[j]})/\tilde{\sigma}_{[j]} \cdot h_{jj}(\lambda)^{1/2}$$

$$= [h_{jj}(\lambda)/(1-h_{jj}(\lambda))]^{1/2} t_{[j]} \quad , \; j = 1,\ldots,n,$$

(see Velleman and Welsch 1981 or Belsley, Kuh and Welsch 1980).

Deleting $q \geq 2$ observations is somewhat more complicated than the case $q = 1$. When $q \geq 2$ it is no longer obvious that equations (3.3) always uniquely determine the $\tilde{y}_j^{(J)}$. This will be true if and only if $(I-H(\lambda))_J$, the submatrix of $I-H(\lambda)$ corresponding to those indices in J, has rank q. For example, when $q = 2$, $J = \{i,j\}$ this condition is equivalent to $(1-h_{ii}(\lambda))(1-h_{jj}(\lambda)) - h_{ij}(\lambda)^2 \neq 0$. Instances where this is not satisfied would seem rare in practice.

Now suppose that one obtains m random samples of q indices each, $J_1,\ldots,J_m$, by sampling with replacement from $\{1,\ldots,n\}$. A bootstrap estimator of the variance-covariance matrix of $\tilde{\beta}$ is provided by

$$\tilde{W} = \Sigma_{r=1}^{m}(\tilde{\beta}^{(J_r)} - \tilde{\beta}^*)(\tilde{\beta}^{(J_r)} - \tilde{\beta}^*)'/(m-1) \qquad (4.8)$$

where $\tilde{\beta}^* = m^{-1}\Sigma_{r=1}^{m}\tilde{\beta}^{(J_r)}$. If the matrices $(I-H(\lambda))_{J_r}$ all have rank q, W can be computed using equations (3.2)-(3.3) and its elements can then be used to obtain bootstrap analogs of the jackknife confidence intervals (4.3). A similar approach when all possible subsets of size q are used leads to the development of grouped jackknife interval estimates of β (see Efron 1982, Chapter 2).

To conclude note that when $\lambda = 0$ Theorems 3.1 - 3.2 can be used to establish "leave-q-out" identities such as equation (7) of Draper and John (1981). It is, therefore, possible to generalize leave-q-out diagnostics such as those discussed in Gentleman and Wilk (1975a, b) and Draper and John (1978, 1981) to the case of penalized least squares estimation.

5. EXAMPLES

In this section the application of results in Sections 3 and 4 to the special cases of smoothing splines and ridge regression will be illustrated.

5.1 Smoothing Splines

Suppose η is a smooth response function and that $\eta_j = \eta(t_j)$, $0 \leq t_1 < \ldots < t_n \leq 1$, in model (1.1). For $n \geq m$ the smoothing spline estimator of η, denoted by $\tilde{\eta}$, is obtained by minimizing

$$\Sigma_{j=1}^{n}(y_j - f(t_j))^2 + \lambda\int_0^1 f^{(m)}(t)^2 dt \qquad (5.1)$$

over all functions f having m-1 absolutely continuous derivatives and a square integrable mth derivative. Schoenberg (1964) proposed this type of nonparametric estimator for η and showed that $\tilde{\eta}$ was a spline function of order 2m with knots at the t_j. General discussions of smoothing splines can be found in Wahba (1977), Wegman and Wright (1983) and Eubank (1983).

Demmler and Reinsch (1975) (see also Speckman 1983) develop a basis for spline smoothing which consists of functions $x_1,\ldots,x_n$ and constants $0 = q_1 = \ldots = q_m < q_{m+1} < \ldots < q_n$ which satisfy

$$\Sigma_{r=1}^{n} x_i(t_r)x_j(t_r) = \delta_{ij} \qquad (5.2)$$

and

$$\int_0^1 x_i^{(m)}(t)x_j^{(m)}(t)dt = q_j\delta_{ij}\ , \qquad (5.3)$$

where δ_{ij} is the Kronecker delta. They show that the minimizer of criterion (5.1) is necessarily of the form

$$f(t) = \Sigma_{j=1}^{n}\beta_j x_j(t) \quad ; \qquad (5.4)$$

hence, it sufficies to minimize criterion (5.1) over functions of this type. Substituting f(t) from (5.4) into (5.1) and invoking the relationships in equation (5.3) gives the equivalent criterion

$$\min_{\beta}\{\Sigma_{j=1}^{n}(y_j-\Sigma_{r=1}^{n}\beta_r x_r(t_j))^2 + \lambda\Sigma_{j=1}^{n}\beta_j^2 q_j\} \ . \qquad (5.5)$$

Comparison with (1.2) reveals this to be a special case of penalized least squares estimation with $p = n$, $x_j' = (x_1(t_j),\ldots,x_n(t_j))$ and $Q = \text{diag}\ (q_1,\ldots,q_n)$. Therefore,

$$\tilde{\beta} = D(\lambda)X'y \qquad (5.6)$$

where $D(\lambda) = \text{diag}((1 + \lambda q_1)^{-1},\ldots,(1 + \lambda q_n)^{-1})$.

The hat matrix corresponding to the estimator (5.6) is $H(\lambda) = XD(\lambda)\hat{X}'$; moreover, since $X'X = I$ the eigenvalues of $(X'X)^{-1}Q$ are simply the q_j. Applying Corollary 1 of Section 2 the following bounds are obtained for $h_{ii}(\lambda)$:

$$h_{ii}(\infty) + (1 + \lambda q_n)^{-1}\Sigma_{r=m+1}^{n}x_r(t_i)^2 \leq h_{ii}(\lambda) \leq h_{ii}(\infty)$$

$$+ (1 + \lambda q_{m+1})^{-1}\ \Sigma_{r=m+1}^{n}x_r(t_i)^2 \ , \qquad (5.7)$$

where $h_{ii}(\infty) = \Sigma_{r=1}^{m}x_r(t_i)^2$. It follows from Demmler and Reinsch (1975) that $h_{ii}(\infty)$ is the ith leverage value for regression on polynomials of order m. Equation (5.7) therefore establishes a connection between the leverage values for spline smoothing and those for polynomial regression. These results generalize to multivariate "Thin Plate" or Laplacian smoothing splines (e.g., Wahba 1981; Wahba and Wendleberger 1980; and Wendelberger 1981) where the $h_{ii}(\lambda)$ may be partilarly useful in the detection of sensitive points in the design.

To illustrate the behaviour of some of the diagnostic and

inferential methods proposed in Section 4, a small scale simulation was conducted. Data sets were generated from model (1.1) with

$$\eta_i = \eta(t_i) = 4.26\{\exp(-3.25t_i)-4\exp(-6.5t_i)+3\exp(-9.75t_i)\},$$

$$t_i = (i-1)/n, \qquad n = 80 ,$$

and normal errors with σ values of .05, .1, .2 and .4. This function is a rescaled version of one studied by Wahba and Wold (1975). The basic experiment was replicated r = 50 times (i.e., 50 data sets of size 80) with each replicate being "treated" by all four values of σ. A cubic smoothing spline (m = 2) was fitted to each data set with λ selected via GCV.

Approximate 95% jackknife confidence intervals for the η_i, centered at $\hat{\eta}_i$, were computed by taking $a_i' = (x_1(t_i),\ldots,x_n(t_i))$ in equation (4.3). The proportion of times the true function value was contained in its interval estimate was recorded along with the value of $\tilde{\sigma}^2$ and the proportion of times $|t_{[j]}|$ exceeded the 5% (two-tailed) critical value for the Student t distribution. Summary statistics for the simulation are given in Table 1. A typical example of these results, for $\sigma = .1$, appears in Figure 1.

[Insert Table 1, Figure 1]

The empirical confidence levels in Table 1 are somewhat lower than might be desired. However, by using 99% rather than 95% intervals, confidence levels in excess of 94% were obtained in all cases. This is typical of simulations performed with other function types and other configurations for the values of r, n, and σ. These results will appear elsewhere. As illustrated in Table 1, the

Student's t approximation to $t_{[j]}$ and the estimator $\tilde{\sigma}^2$ performed well.

5.2 Ridge Regression

Ridge regression estimators (Hoerl and Kennard 1970; Marquardt 1970) are solutions to the criterion function (1.2) when (i) only the nonconstant predictor variables from model (1.1) are included in X, (ii) the predictor variables are standardized so that X'X is in correlation form, and (iii) Q = I. Much controversy persists over automated selection of λ, the effect of standardization on ridge estimation, and the assumptions underlying the validity of the ridge estimator (e.g., Draper and Van Nostrand 1979; Smith and Campbell 1980, with discussion). In order to demonstrate the application of the results of Section 2-4, assume that for a specific regression analysis the criticisms noted above are satisfactorily answered and that a ridge regression analysis is deemed appropriate.

Ridge regression diagnostics can be obtained from the results of Sections 2-4 under the conditions stated above; however, the efficient computational expressions for deleted estimators (i.e., $\tilde{\beta}^{[j]}$ and $\tilde{\beta}^{(J)}$) and deleted residuals (i.e., $e_{[j]}$) are exact only if the reduced X matrix is not restandardized when rows are deleted. Hinkley (1977a) noted a similar restriction when he cautioned against obtaining (least squares) jackknife estimates of the constant term of a regression model using centered predictor variables. Since the major benefits of centering and standardization cited by Marquardt (1980) are essentially maintained when one (or a small number) of the rows of the standardized X matrix is (are) deleted,

only the original matrix of predictor variables is standardized in the following example.

Gunst and Mason (1980, Appendix A) contains a data set on the gross national product (GNP) of 49 countries of the world along with the six additional socioeconomic indices: an infant death rate (INFD), a physician/population ratio (PHYS), population density (DENS), population density measured in terms of agricultural land area (AGDS), a literacy measure (LIT), and an index of higher education (HIED). Table 2 displays regression diagnostics for the fit of ℓn(GNP) by the six socioeconomic indices.

[Insert Table 2]

The relatively small value of $\lambda(0.08)$ which was chosen for this illustration has little effect on the bounds for ridge leverage values given by Corollary 2 since $\ell_6/(\ell_6+0.08)=0.97$. With the exception of Malta, least squares leverage values which exceed $2(p+1)/n = 0.286$ are also large with the ridge estimator using the analogous bound $2(tr[H(\lambda)]+1)/n = 0.271$. Although the ridge DFFITS values appear to be slightly more uniform than those of least squares (e.g., none of the former are greater than 1.0 in magnitude), four of the five observations which exceed $2\{(p+1)/n\}^{1/2}=0.756$ for least squares also exceed $2\{(tr[H(\lambda)]+1)/n\}^{1/2}=0.736$ for ridge regression--Malta is again the exception--and a similar comment can be made about the $t_{[j]}$.

Malta is obviously affecting the two estimation procedures differently. It has high leverage and is influential on the least squares fit but has neither high leverage nor an influential impact on the ridge regression fit. A scatterplot of DENS and AGDS reveals that Malta lies well off the concentrated linear scatter (r = 0.97)

between these two variates. Thus by lessening the effect of the strong pairwise correlation between DENS and AGDS on the estimation of the regression coefficients, the ridge estimator is also lessening the influence of Malta on the fit. Although the other least squares and ridge diagnostics identify equally important characteristics of this data set, comparison of the two sets of diagnostics has provided important insight about Malta which might have gone unappreciated had only the least squares diagnostics been examined.

Table 3 displays least squares, ridge ($\lambda = .08$), and jackknifed ridge ($\tilde{b}$) coefficient estimates and confidence intervals. The purpose of presenting the ridge and jackknifed ridge estimates is to highlight typical characteristics of these estimators, not to draw definitive conclusions relative to this data set. Note in particular that, while similar, the ridge and jackknifed ridge estimates are somewhat different. In addition, both of these latter two estimators produce jackknife confidence intervals (using expressions (4.3)) which are shorter than least squares. In view of the simulation results in Section 5.1, it might be advisable to adjust these confidence intervals (not done here) by using a larger Student t critical point. If one uses 99% nominal coverage, the ridge confidence interval for the coefficient of DENS includes the origin.

Obviously a more complete analysis of this data set is needed in order to resolve questions which remain about influential observations and the significance of the predictor variables. Any thorough

analysis must incorporate prior knowledge about the regression coefficients and information concerning the intended use of the conclusions which are to be drawn from the fitted model. These topics are beyond the scope of this paper; nevertheless, this example illustrates some important characteristics of penalized least squares diagnostics and approximate inference procedures.

6. CONCLUDING REMARKS

The results of this paper generalize least squares regression diagnostics and certain approximate inference procedures to a class of (quadratic) penalized least squares estimators for linear models. Theorems 3.1 and 3.2 produce expressions for deleted estimators and residuals which provide exact, computationally efficient, calculation of quantities such as pseudo values and Studentized residuals. These results have wide application, two specific illustrations being nonparametric estimation with smoothing splines and ridge regression.

Much research remains to be conducted regarding the properties and usage of the procedures proposed in this paper. For example, the jackknife confidence intervals do not achieve the nominal confidence level, although they are well-known to be insensitive to a variety of unimodal error distributions. Corrections for the jackknife such as those proposed in Hinkley (1977b, 1978) may alleviate coverage difficulties and the behavior of jackknife intervals under nonnormal errors merits further investigation. Likewise, the sensitivity of jackknife confidence intervals to the choice of λ warrants

further study. For instance, in the ridge regression example increasing λ from 0.08 to 0.20 decreases the estimated standard errors of the individual coefficients between 5 percent (HIED) and 50 percent (AGDS). On the other hand, the Studentized residuals and the estimator of σ^2 performed well in the simulation in Section 5.1. Similarly, the ridge regression diagnostics highlighted an important characteristic of the presence of Malta which could have been overlooked if only the least squares diagnostics were examined.

REFERENCES

BECKMAN, R. J. and TRUSSELL, H. J. (1974), "The Distribution of an Arbitrary Studentized Residual and the Effects of Updating in Multiple Regression," Journal of the American Statistical Association, 69, 199-201.

BELSLEY, D. A., KUH, E. and WELSCH, R. E. (1980), Regression Diagnostics, New York: John Wiley and Sons, Inc.

CRAVEN, P. and WAHBA, G. (1979), "Smoothing Noisy Data with Spline Functions," Numerische Mathematik, 31, 377-403.

DEMMLER, A. and REINSCH, C. (1975), "Oscillation Matrices with Spline Smoothing," Numerische Mathematik, 24, 375-382.

DRAPER, N. and JOHN, J. A. (1978), "On Testing for Two Outliers or One Outlier in Two-Way Tables," Technometrics, 20, 69-78.

DRAPER, N. R. and JOHN, J. A. (1981), "Influential Observations and Outl ers in Regression," Technometrics, 23, 21-26.

DRAPER, N. R. and VAN NOSTRAND, R. C. (1979), "Ridge Regression and James-Stein Estimation: Review and Comments," Technometrics, 21, 451-466.

EFRON, B. (1982), The Jackknife, the Bootstrap, and Other Resampling Plans, Philadelphia: Society for Industrial and Applied Mathematics, Monograph No. 38.

EUBANK, R. L. (1983), "Approximate Regression Models and Splines," Communications in Statistics: Statistical Reviews (to appear).

GENTLEMAN, J. F. and WILK, M. B. (1975a), "Detecting Outliers in a Two-Way Table: I. Statistical Behavior of Residuals," Technometrics, 17, 1-14.

GENTLEMAN, J. F. and WILK, M. B. (1975b), "Detecting Outliers. II. Supplementing the Direct Analysis of Residuals," Biometrics, 31, 387-410.

GOLUB, G. H., HEATH, M., and WAHBA, G. (1979), "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter," Technometrics, 21, 215-223.

GUNST, R. F. and MASON, R. L. (1980), Regression Analysis and Its Application: A Data-Oriented Approach, New York: Marcel-Dekker, Inc.

HINKLEY, D. V. (1977a), "Jackknifing in Unbalanced Situations," Technometrics, 19, 285-292.

HINKLEY, D. V. (1977b), "Jackknife Confidence Limits Using Student t Approximations," Biometrika, 64, 21-28.

HINKLEY, D. V. (1978), "Improving the Jackknife with Special Reference to Correlation Estimation," Biometrika, 65, 13-21.

HOAGLIN, D. C. and WELSCH, R. E. (1978), "The Hat Matrix in Regression and ANOVA," American Statistician, 32, 17-22.

HOERL, A. E. and KENNARD, R. W. (1970), "Ridge Regression: Biased Estimation for Nonorthogonal Problems," Technometrics, 12, 55-67.

KSHIRSAGAR, A. M. (1972), Multivariate Analysis, New York: Marcel Dekker, Inc.

MARQUARDT, D. W. (1970), "Generalized Inverses, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation," Technometrics, 12, 591-612.

MARQUARDT, D. W. (1980). "You Should Standardize the Predictor Variables in Your Regression Models," Journal of the American Statistical Association, 75, 87-91.

MILLER, R. G. (1974), "An Unbalanced Jackknife," Annals of Statistics, 2, 880-891.

PREGIBON, D. (1981), "Logistic Regression Diagnostics," The Annals of Statistics, 9, 705-724.

SCHOENBERG, I. J. (1964), "Spline Functions and the Problem of Graduation," Proceedings of the National Academy of Sciences (USA), 52, 947-950.

SMITH, G. and CAMPBELL F. (1980), "A Critique of Some Ridge Regression Methods," Journal of the American Statistical Association, 75, 74-81.

SPECKMAN, P. (1983), "The Asymptotic Integrated Error for Smoothing Noisy Data by Splines," Numerische Mathematik (to appear).

VELLEMAN, P. and WELSCH, R. E. (1981), "Efficient Computing of Regression Diagnostics," The American Statistician, 35, 234-242.

WAHBA, G. (1977), "A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for Solving Them," In Applications of Statistics (P. R. Krishnaiah, ed.) 507-523. Amsterdam: North Holland.

WAHBA, G. (1981), "Numerical Experiments with the Thin Plate Histospline," Communications in Statistics, A10, 2475-2514.

WAHBA, G. (1983), "Bayesian Confidence Intervals for the Cross-Validated Smoothing Spline," Journal of the Royal Statistical Society, Series B, 45, 133-150.

WAHBA, G. and WENDELBERGER, J. (1980), "Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross Validation," Monthly Weather Review, 108, 1122-1143.

WAHBA, G. and WOLD, S. (1975), "A Completely Automatic French Curve: Fitting Spline Functions by Cross Validation," Communications in Statistics, A4, 1-17.

WEGMAN, E. J. and WRIGHT, I. W. (1983), "Splines in Statistics," Journal of the American Statistical Association, 78, 351-365.

WENDELBERGER, J. (1981), "The Computation of Laplacian Smoothing Splines with Examples," Technical Report No. 648, Department of Statistics, University of Wisconsin-Madison.

TABLE AND FIGURE TITLES

Tables

1. Summary Statistics for the Simulation
2. Regression Diagnostics for GNP Data, Selected Observations
3. Coefficient Estimates and Nominal 95% (Individual) Confidence Intervals

Figure

1. Typical Jackknife Confidence Intervals, Spline Simulation

TABLE 1. Summary Statistics for the Simulation

	Empirical Confidence Levels		Empirical Significance Levels		Estimated Variance	
σ	Average	Std. Error	Average	Std. Error	Avg.	MSE
.05	.8838	.0084	.0508	.0025	.0023	2×10^{-7}
.10	.8868	.0087	.0510	.0024	.0091	3×10^{-6}
.20	.8863	.0102	.0493	.0021	.0366	5×10^{-5}
.40	.8843	.0149	.0490	.0023	.1490	6×10^{-4}

TABLE 2. Regression Diagnostics for GNP Data, Selected Observations

	Least Squares			Ridge (λ=.08)		
Obsn.	h_{jj}	$t_{[j]}$	$DFFITS_j$	$h_{jj}(.08)$	$t_{[j]}$	$DFFITS_j$
BARBADOS	.238	-2.026	-1.131	.137	-1.929	-.769
CANADA	.042	2.011	.419	.039	2.111	.423
HONG KONG	.511	-.107	-.109	.471	-.138	-.130
INDIA	.558	1.337	1.502	.507	.903	.917
JAPAN	.049	-2.799	-.633	.046	-2.743	-.602
LUXEMBOURG	.084	2.356	.713	.077	2.391	.690
MALTA	.688	1.506	2.236	.262	.426	.254
SINGAPORE	.632	.562	.736	.516	.632	.653
TAIWAN	.178	-2.401	-1.119	.129	-2.475	-.953
U.S.	.490	.804	.787	.447	.951	.855

TABLE 3. Coefficient Estimates and Nominal 95% (Individual) Confidence Intervals

Predictor Variable	Least Squares Estimates	Ridge Regression (λ = .08)	Jackknifed Ridge
	(a) Coefficient Estimates		
INFD	-1.870	-1.772	-1.695
PHYS	.171	- .125	.113
DENS	-1.094	- .410	- .606
AGDS	.862	.151	.453
LIT	2.298	1.985	2.163
HIED	1.454	1.411	1.662
	(b) Confidence Intervals		
INFD	(-3.012,- .729)	(-2.218,-1.326)	(-2.142,-1.250)
PHYS	(-1.192, 1.535)	(- .524, .274)	(- .286, .512)
DENS	(-4.718, 2.530)	(- .767,- .053)	(- .963,- .249)
AGDS	(-2.738, 4.462)	(- .188, .490)	(.114, .792)
LIT	(.748 3.848)	(1.408, 2.562)	(1.586, 2.740)
HIED	(.528, 2.380)	(.994, 1.828)	(1.245, 2.079)